

POVERTY TARGETING AND MEASUREMENT TOOLS IN MICROFINANCE

Progress out of Poverty Index and the Poverty Assessment
Tool

October 2010

FORD FOUNDATION

 CGAP



Acknowledgments: This report was written with the support of the European Commission, the Ford Foundation, and CGAP. CGAP coordinated the research phase and participated in two field reviews. The report follows a series of consumer reviews of social performance tools used in microfinance (<http://www.sptf.info/page/user-reviews-of-sp-tools>). The authors would like to thank Nigel Biggar from Grameen Foundation, Brian Beard, Margaret Richards, and Anthony Leegwater from the IRIS Center at the University of Maryland, Mark Schreiner from Microfinance Risk Management, L.L.C., and the participating MFIs for their outstanding cooperation in the review process. The openness of all parties and the collaborative stance are to be noted.

Table of Contents

1. Report Goals.....	4
2. What are poverty scorecards? Two examples: PPI and PAT	5
Part I: Accuracy Analysis of PPI and PAT	8
3. Dimensions of Tool Accuracy Reviewed	8
How are the scorecards put together?	8
What are the key differences between the two tools?	10
Tool accuracy.....	12
Determinants of accuracy	15
What kinds of questions do scorecards answers?	20
Can the scorecards be used for targeting?	21
Can the scorecards be used for impact evaluation?	22
Part II: User review, what do MFIs say?	24
4. Methodology	24
5. User and External Review of PPI.....	25
6. User and External Review of PAT	30
7. Summary table.....	34
8. Recommendations.....	35
9. Annexes.....	39
10. References.....	56

1. Report Goals

The area of poverty targeting and measurement in microfinance is particularly of interest to microfinance stakeholders, amidst increasing criticism that the industry is straying from its original mission of poverty alleviation. A small percentage of microfinance institutions (MFIs) have developed their own scorecards or measurement tools to track client poverty, while a larger percentage appear to use household income or household expenditure analysis and geographic targeting¹. This report aims to provide objective reviews of two standardized tools that allow MFIs to measure their poverty outreach and targeting of clients², focusing on two criteria: accuracy and user satisfaction. While the report is an initial exploration of this subject and more research is needed, the tools, the Poverty Assessment Tool (PAT) and the Progress out of Poverty Index (PPI), hold promise of being cost-effective ways to target and track outreach. Challenges in the way the tools are developed and implemented remain.

Part I of this report provides an overview of how the tools are constructed; explores different dimensions, such as accuracy and implementation, of the tools; and gives specific recommendations to further improve scorecards in microfinance. **Part II**—the user reviews—summarizes the feedback given by implementers on a set of six indicators.³ The report concludes with **recommendations** on further analytical and operational piloting of scorecards in microfinance.

The research is based on an analysis of the statistical methods used to construct and implement the tools, review of written documentation available on the tools' websites,⁴ discussions with tool developers, and discussions and field visits to **7 MFIs that have used the tools**. The research process benefited from an advisory committee comprised of Frank De Giovanni, Ford Foundation; Xavier Reille and Antonique Koning, both from CGAP; Ghazala Mansuri, Senior Economist at the World Bank; and Gabriel Solorzano from MixCapital (former Managing Director of Banex Nicaragua). The main contributors to this report are Stephen Boucher (Economist, Associate Professor at UC Davis), Reuben Summerlin (Microfinance Specialist, ACT), and Meritxell Martinez (Microfinance Analyst, CGAP).

This report is for managers of MFIs, staff involved in monitoring and evaluation, donors and investors concerned with poverty outreach of MFIs, and policy makers. An attempt is made to minimize the technical language; however, some sections will make reference to basic statistical concepts.

¹ MIX Social Performance database and CGAP Survey on Targeting tools

² Note that only the Progress out of Poverty Index has been used for targeting purposes

³ Cost, Duration of Implementation, Usefulness, Robustness of Results, Ease of Use, and Overall Satisfaction

⁴ The websites for PAT and PPI are <http://www.povertytools.org/> and <http://progressoutofpoverty.org/>, respectively.

2. What are poverty scorecards? Two examples: PPI and PAT

For purposes of this report, a scorecard is a poverty measurement tool (PMT) that allows an MFI to answer the following general questions: **What percentage of my clients is poor? How does that percentage change over time?**

The scorecard or short survey is the most obvious component of the PMT and is applied to individual clients as a survey.⁵ PPI and PAT⁶ are two examples of scorecards that are gaining ground in the microfinance industry and beyond, since these tools can be used by other social enterprises. The development of these surveys started at the beginning of 2000, supported by different donors interested in quantifying microfinance clients' absolute poverty levels and benchmarking those with international poverty lines. These surveys collect information on key indicators, which are the household-level variables that help predict the poverty status of the household (either poor or nonpoor).

	PROGRESS OUT OF POVERTY	POVERTY ASSESSMENT TOOL
QUESTIONS IT ANSWERS	What percentage of my clients is poor? How does that percentage change over time?	What percentage of my clients is poor? How does that percentage change over time?
MAIN TOOL SPONSOR⁷	Grameen Foundation	USAID
TOOL DEVELOPER (technical)	Microfinance Risk Management, L.L.C. (Mark Schreiner) http://www.microfinance.com/	IRIS Center University of Maryland http://www.povertytools.org/
NO. OF COUNTRIES WHERE IT IS AVAILABLE	34	30
OUTPUT (according to developer)	An estimation of a group's poverty rate at a point in time and an estimation of changes in a group's poverty rate between two points in time	An estimation of the poverty outreach of an organization, as a percentage of its client population that is below one or more poverty lines.

In the case of PPI, the household's data are used to predict the probability that the household is poor. The poverty incidence of the MFI is then estimated by taking the average of these predicted

⁵ Throughout this report, we will use the terms "survey", "poverty measurement tool—PMT", and "scorecard" interchangeably. We note that the term "scorecard", however, is more appropriate for PPI's survey because the points or "scores" for each question are normally shown on the survey form. In contrast, PAT uses a more traditional survey that does not explicitly show the relative, quantitative contribution of each question to predicted household expenditures or the probability that the household is poor.

⁶ Examples of the latest version of the Peru surveys associated with each tool are provided in annexes 1 and 2.

⁷ In terms of funding committed to develop PPI at national level, there have been 48 new or updated simple poverty scorecards/PPIs. Grameen Foundation funded 24 (half) of these with funding from the Ford Foundation, Microcredit Summit Campaign funded 5, CGAP funded 13, Trickle Up funded 1, UNRWA 3, HIVOS 1, and ICCO 1. There are 31 PPIs available at Grameen Foundation website: <http://progressoutofpoverty.org/ppi-country>.

probabilities across all surveyed households. In the case of PAT, the household's data are used to predict the per capita level of expenditures (or income) of the household. Each surveyed household is then classified as poor or nonpoor by comparing its predicted expenditure level to a particular poverty line. The poverty incidence of the MFI is then estimated by calculating the percentage of surveyed households that are predicted to be poor.

Through the indicators that are selected and the weights assigned to each indicator, the tools capture underlying relationships between household characteristics and poverty. Because these relationships differ across countries, **the tools are country specific**. Currently, the developers of both tools provide a single tool for each country. Thus, for example, **there is one and only one** PAT for Peru and one and only one PPI for Peru.⁸ Because the tools use different algorithms, an MFI could get a different estimate of poverty incidence depending on which tool it uses. An interesting exercise for future research would be to conduct a head-to-head comparison (i.e., apply the tools to the same data set or to the same clients of an MFI) of the estimates of poverty incidence of the two tools.

It is important to note that, while the survey (and the indicators it includes) is the most observable component of a PMT, a PMT consists of a number of additional components, each of which is fundamental for the ultimate accuracy of the tool.

Additional components associated with construction of the tool by the tool developers include the following:

- A methodology for choosing the indicators that appear in the survey.
- A statistical algorithm that assigns weights to each indicator. The weights define the relative contribution of each indicator to the predicted expenditures of the household (PAT) or the probability that the household falls below a particular poverty line (PPI).

Additional components associated with implementation of the tool by MFI's include the following:

- Guidance on appropriate and statistically sound techniques to select the sample of clients to whom the survey is applied
- A data entry program permitting transfer of the survey responses to the MFI's information systems
- A software program that uses the collected data to generate the poverty estimate
- Training material to teach the MFI how to implement the tool and interpret results
- Technical support for each step of tool implementation and use

Each of these requirements has different implications when implementing a scorecard at the MFI level as observed during the field visits. MFIs with more access to training materials, better

⁸ It is, of course, possible that within a given country the relationship between poverty and household indicators varies across different subgroups, such as those defined by geographic location or ethnic group. As such, it is sensible to consider creating multiple, group-specific versions of a given tool within a country. While creating, for example, state-specific PPI's in India would, in theory, increase accuracy, the gain would come at a significant cost. At this point, the developers of both tools conclude that the costs of developing multiple tool versions per country outweigh the likely gain in accuracy. Schreiner (2006) provides evidence that accuracy would not be significantly increased by instituting separate rural and urban scorecards for India.

technical support, or more developed MISs, for instance, would have a smoother implementation of the scorecards.

Currently, there are **PPIs for 34 countries and PATs for 30 countries**. These countries, and the year of their most recent update, are listed in Annex 3.⁹ At the MFI level it is not possible to know how many MFIs are currently using these tools. However, MIX social performance data gives an indication: 22 MFIs reported having piloted or used PPI, and 15 provided their results to MIX. For PAT in 2009, 40 MFIs and enterprise development providers have reported their results to the USAID Microenterprise Results Reporting (MRR) system.

⁹ A PPI is currently being developed for 7 additional countries. PATs will be released in the next few months for 6 additional countries, with others in longer term development

Part I: Accuracy Analysis of PPI and PAT

3. Dimensions of Tool Accuracy Reviewed

How are the scorecards put together?

This section provides a general description of the two most important pre-implementation components of the two PMTs: (1) The methodology used to select indicators for inclusion in the survey and (2) the methodology used to construct the weights associated with each indicator.

Identifying a Nationally Representative Household Income/Expenditure Survey

The starting point for both tools is to identify a high-quality, nationally representative survey that measures household income or expenditures, such as the World Bank's Living Standard Measurement Survey (LSMS). These surveys are designed to measure national poverty levels and thus tend to have detailed and high-quality data to measure household expenditures or income. They also include a large number of household-level variables, such as asset holdings, sociodemographic characteristics of household members, and variables describing housing quality. These variables, which often number more than 500, represent the potential set of indicators from which the tool developers choose indicators for inclusion in the poverty tool.

Because the national survey serves as the basic input into the construction of the poverty tool, **the quality of the PMT is ultimately limited by the quality of the national survey data that are available**. The lower the quality of the national data (i.e., the greater the error with which it is measured), the lower the quality of the scorecard.¹⁰

It is also important to note that the relationship between specific household characteristics (indicators) and poverty changes over time. As a result, even when the national household data are collected carefully and with a high degree of accuracy, the older the national data upon which a scorecard is based, the less accurate the scorecard will be. During the field interviews the MFIs interviewed expressed their feeling that some of the indicators were not relevant today. We return to this important point later in this report. For now, we note that an important **recommendation** of this report is to conduct more work to examine how quickly the accuracy of a poverty tool deteriorates over time when it is not updated with a new national data set.

Indicator Selection and Assignment of Weights

The next steps are to select the indicators that will ultimately appear on the survey and to assign each indicator a weight representing the relative contribution of that indicator to the household's predicted poverty status. While these appear to be two separate steps, they in fact occur simultaneously. Both tools use a regression framework to simultaneously select the indicators and

¹⁰ The developers of both tools screen national data sets to evaluate their quality. If the available data set exhibits significant problems due, for example, to improper or unclear sampling methodology or lack of recent updating, the data set will be rejected. PAT developers also screen out existing data sets that measure expenditures using inappropriate methodology. In some cases where existing data are problematic, PAT developers have carried out their own household expenditure survey to construct the tool.

assign weights. In general, regression is just a statistical means of analyzing the relationship between a dependent variable (the outcome that we want to explain) and independent variables (the factors that explain differences in the dependent variable).

First, the dependent variable is constructed for each household in the national data set. The dependent variable is based on either income per capita or expenditure per capita for each household, depending on which type of data was collected. In general, expenditure data are preferred because they are easier to collect and are a better indicator of household welfare than income.¹¹ Different variables from the household survey are then selected as independent variables in the regression model. For a given set of independent variables, the regression model will examine the correlations between the independent variables and the dependent variable and generate a regression coefficient for each independent variable. This coefficient summarizes the direction (do household expenditures tend to increase or decrease when the independent variable increases?) and magnitude (is the increase or decrease small or large?) of the correlation. The regression coefficients, in turn, are then converted into the weights used in the scorecard.

For both tools, a combination of **statistical** and **subjective** criteria is used to select the final set of indicators. The statistical criterion consists of examining how the predictive power of the regression model changes when a specific indicator is included. In the case of PPI, the final set of indicators should maximize the probability that households' poverty statuses are correctly predicted. In the case of PAT, the selection of final indicators is based on their ability to explain a higher percentage of variation in household expenditures as well as balancing accuracy at the household level (are truly poor households correctly classified as poor and truly nonpoor households classified as nonpoor?) with accuracy at the aggregate/group level (on average, does the tool deliver the true poverty incidence of a given group of households, such as the clients of an MFI?).

Balanced against this statistical criterion are two practical issues. First, the tool developers may include some variables that are likely to change over time (ownership of a small appliance, for instance) at the expense of other indicators that may have higher predictive power but are less likely to change (education level of household head).¹² This first issue is more directly relevant to PPI since, from its inception, PPI developers identified measuring the change in an MFI's poverty rate over time as an explicit objective. In addition, as will be further discussed later, the PPI scorecard asks a smaller number of questions than does the PAT survey. Second, both tools seek to include indicators that can be measured with a high degree of accuracy while imposing minimal cost on the MFI and the respondent. Thus some indicators that have relatively high predictive power may be rejected because they are either difficult to recall (quantity of meat consumed in the last week) or they make the respondent uncomfortable (how much money do you have in your bank account?). This subjective cost-benefit calculation is done by the tool developers on a

¹¹ Income data are notoriously difficult to collect in developing countries because of the prevalence of self-employment (which requires the reconstruction of revenue and costs) and nonmarketed production. Expenditures (consumption) are a better indicator of welfare because consumption need not fluctuate as much as income because of households' ability to "smooth" consumption via savings, insurance, and borrowing.

¹² This discussion of the need to include potentially time varying indicators to measure changes in poverty incidence is based on the individual PPI country reports written by Mark Schreiner and available at <http://www.microfinance.com/>.

country-by-country basis. In some cases, the tool developers consult with MFIs in the country to make the final determination of which indicators are selected.¹³ During the field visits, MFIs consulted in the subjective process of selecting the indicators appeared more pleased with the final list of questions than those who were not consulted. Moreover, some MFIs expressed their concern of having indicators that had weights that were “too dominant” in the final score (in the case of PPI). We note that an important **recommendation** of this report is to encourage more consultations in the process of selecting indicators (with MFIs networks, MFIs, investors, etc.). Users’ understanding on how the tool’s predictive power depends on the whole set of indicators and not in specific ones is very low.

What are the key differences between the two tools?

While the general approach to selecting indicators and constructing weights is similar across the two tools, there are several notable differences.

The **first** difference is that they define a **different target number of indicators** to be included. As seen in Annexes 1 and 2, PPI uses a smaller number of indicators, typically only 10, while PAT includes approximately 15 to 20 indicators. In contrast to PPI, PAT typically requires the interviewer to fill out a household roster,¹⁴ including the age, gender, relationship to head of household, and education of each family member. As a result, for a household of 5 individuals, this effectively implies an additional 20 questions. Again in contrast to PPI, PAT also includes at least three regional variables plus an indicator of whether the household lives in a rural or urban area.

Two points are worth mentioning about these differences in variables. First, if the relationship between poverty and the indicators varies significantly across regions, then the inclusion of a regional indicator (as PAT does) would likely improve the accuracy of the tool. One must be cautious, however, when including a regional indicator for two reasons. If the national data set has only a small number of observations per region, then the regional indicators may be quite imprecise (have high standard errors) even though they are large in magnitude. Second, if an MFI operates in multiple regions, there may be confusion and errors in recording the regional indicator. The inclusion of additional variables, and particularly the household roster, is likely to make PAT more time-consuming and thus more costly to implement. A **recommendation** of this report is to evaluate the tradeoff between the cost versus the potential gain in accuracy associated with the inclusion of these variables as well as regional indicators.¹⁵

The **second** primary difference is that the tools rely on **different statistical methods** to select the indicators and generate the weights associated with each indicator.¹⁶ PPI uses a Logit regression model to select the indicators that are included in the scorecard. In a Logit model, the dependent variable is a binary variable that takes value 1 if the household is nonpoor and 0 if the household is

¹³ In the initial development stage of PAT, the developers carried out extensive fieldwork in a number of countries in order to identify the most difficult and problematic types of questions. For a summary of this work in Peru, for example, see IRIS (2006). This work has been supplemented by subsequent experience and aggregated user feedback.

¹⁴ According to the PAT developer, “fidelity to the manner in which the question was asked in the original questionnaire improves the accuracy of the reported data.”

¹⁵ As the second part of this report indicates, implementing PPI takes 1–20 minutes and implementing PAT 10–40 minutes per client.

¹⁶ Details about the methodological differences are noted in Annex 7.

poor. The developer¹⁷ creates this variable by determining whether or not the household's total expenditures per capita are greater or less than the chosen poverty line. If greater, the household is classified nonpoor, and if less it is classified as poor. In constructing the PPI scorecard, the developer begins by randomly separating the national household survey data into three separate subsamples: (1) construction sample, (2) calibration sample, and (3) validation sample. The construction sample is used to run the Logit regression in order to select the indicators and generate the weights. The "points" that appear on the scorecard (see PPI scorecard in Annex 2), in turn, are transformations of the coefficient associated with each indicator (independent variable) in the Logit regression. Given the way the transformation is done, the points always add up to a number between 0 and 100. The calibration sample is then used to generate the "lookup table," which gives the probability that a household is poor based on its point total. Annex 4 provides the lookup table associated with the national poverty line in Peru. The technique is straightforward; the probability that a household that scored 0–4 points is poor is given by the fraction of households in the calibration sample that would have scored 0–4 points that were poor.

In contrast, **PAT explores a variety of regression models**, including the Probit and Linear Probability models, which are similar to the Logit model used in PPI, as well as OLS and Quantile models that use the continuous household expenditure per capita variable as the dependent variable. When developing PAT, the developers thus compare both across indicators within a given regression model as well as across regression models and then choose the model (and indicators) that generates the most accurate predictions of household poverty. Given both the larger number of indicators and the greater number of models considered (including the Probit), it would appear that PAT delivers greater accuracy in predicting poverty. A direct comparison of the accuracy of the two tools across countries where both tools exist was beyond the scope of this review and would be an important area for future work.

The **third** general difference in the construction of the survey/scorecard is the **degree of information available to the individuals implementing the survey**. In the case of PPI, the weights (or "points") of each indicator are explicit in the scorecard, and as described above, the surveyor can immediately calculate the household's poverty likelihood. Again, with reference to the PPI scorecard and "lookup table" in Annex 4, we can see that the person applying the survey can immediately see which questions and which answers to each question will most strongly affect the predicted likelihood of poverty of the household. A household that has a wood floor instead of a dirt floor, for example, would earn an additional 17 points. If this raises the total score from 20 to 37, then the estimated probability that the household is poor would drop from 75.4% to 38.1%. In contrast, the weights are not explicitly available to the surveyor implementing PAT. As a result the predicted expenditures (or poverty likelihood for some models) of an individual household is *not* known at the time of survey application. There are two means by which this difference in visibility of the weights could have an impact on accuracy. On one hand, if incentives exist for the individual applying the survey (or the MFI itself) to report higher than true poverty levels, then the greater visibility in PPI could make misreporting more likely. On the other hand, greater visibility of the weights in PPI may make PPI more attractive to an MFI precisely because of the greater

¹⁷ Mark Schreiner of Microfinance Risk Management, L.L.C.

transparency of how the indicators contribute to the poverty score. If manipulation appears to be an issue, then a case could be made for auditing PPI scores. In interviews with MFI users, loan officers appeared to remember the questions with a higher weight in the final score very well and disliked them because they “counted too much,” or were too dominant in the final score.¹⁸ At this point, we simply acknowledge the potential for manipulation of PPI scores and **recommend** follow-up analysis of this point.¹⁹

For now, we note that differences in accuracy across the two tools due to the different underlying statistical methodologies are likely to be relatively small. Of much greater concern to accuracy is the **potential for improper implementation** of the two tools. Based on our field visits, it would appear that tool implementation is likely to be the largest concern for accuracy, dominating any differences in the statistical algorithms underlying the two models.

Tool accuracy

What do we mean by “accuracy”? Little bias and high precision As stated at the beginning of this report, the primary question that both tools are designed to answer is “What is the percentage of clients that are poor?” This percentage is known as poverty incidence. To answer this question, the MFI applies the tool to a random sample of clients or all of its clients and generates an *estimate* of the MFI’s poverty incidence. MFI managers and donors alike want this estimate to be as good as possible, reflecting as closely as possible the “true” poverty rate of a group of clients.

In general, we desire tools that have as little bias as possible while having a high degree of precision.²⁰

How are bias and precision measured? To summarize the accuracy of the tools, we refer to the documentation provided by and discussion with the tool developers. The following section assumes a level of understanding of technical language and targets monitoring and evaluation officers of MFIs or technical staff. We note here that there is a difference across the two tools in terms of the written information provided about the methodologies used to evaluate bias and precision. Specifically, the country-specific reports provided by the PPI developer provide substantially greater information about how bias, and especially, precision are measured relative to the PAT website.

How accurate are the tools?

PPI. PPI developers provide extensive documentation regarding accuracy. An in-depth report is available for each country’s PPI. The report provides a detailed description of the construction of the scorecard, including a discussion of the national data set upon which the scorecard is based, and

¹⁸ See user feedback in Part II of this report.

¹⁹ As Mark Schreiner pointed out in personal communication, it is worth mentioning two caveats to this discussion. First, while the default method of applying PPI is to display the points on the scorecard, the MFI can choose to use a scorecard that does not include the points. In this case the MFI would calculate poverty incidences later at a central location. Second, even when points are not explicitly given on the scorecard, the enumerator likely knows which responses are associated with higher and lower probabilities that the household is classified as poor. As such, excluding weights on the scorecard does not imply that manipulation does not occur.

²⁰ See Annex 5 for a more in-depth explanation on both concepts applied to microfinance.

a comparison of PPI versus other existing PMTs for that country. In addition, the report carries out and summarizes extensive accuracy analysis. The basic methodology of the accuracy analyses developed by the PPI tool developer is as follows. First, the developer uses one portion of the overall sample to select the indicators and construct the weights. A separate portion of the sample is kept aside (the validation sample) in order to evaluate bias and precision. To do so, the tool is applied to many, for example 1,000, randomly selected samples from this validation sample. Bias is estimated by calculating the average of the poverty incidence across **these 1,000 samples and comparing this average to the true poverty incidence** of the entire validation sample. Precision is measured by examining the degree of variability of poverty incidence estimates **across these 1,000 samples**.²¹

For this report, we limit the discussion to two types of accuracy analyses. First, the PPI reports present—for the different ranges of poverty scores—the difference between the true versus the average poverty incidence estimated by PPI using the “validation” sample described above.²² This difference provides one way of gauging the bias in the tool. If this difference is positive, then the tool underestimates the true poverty incidence and vice-versa if the difference is negative. This type of analysis gives a good indication of the accuracy of PPI when applied to individual households. An example of this type of analysis is provided in Annex 6, which is taken from the PPI report for Bangladesh (Chen and Schreiner 2009). Consider the third row, which focuses on those households from the validation sample that scored 10–14 points on PPI. The “-5.9” figure says that, on average, PPI underestimated the probability that households in this range of scores were poor by 5.9%. This difference—or bias—varies across the possible ranges of scores with some being positive and some being negative. This figure also provides a measure of the precision of PPI. Again refer to the 10–14 point range, and focus on the “3.5” figure (the confidence interval) in the 90 % confidence column. This says that for 90% of samples that we draw, the difference between the estimated and true poverty incidence of households in this range would fall between -9.4 and -2.4 (i.e., $-5.9 - 3.5 = -9.4$ and $-5.9 + 3.5 = -2.4$).

Second, and more directly relevant for MFIs, the developers report **accuracy results for groups of households**. Figure 10 in Annex 6 presents the results of this group accuracy analysis for Bangladesh. Consider, for example, the entries in the eighth row. The results indicate that if we drew samples of 256 households—which is close to the sample size typically implemented by MFIs—then, on average, PPI would underestimate the poverty incidence of the group by 1.6 percentage points. In addition, if we were to draw many different samples of size 256, 90% of the estimates generated by PPI would fall between -5.6 below to 3.4 percentage points above the true poverty incidence.

PAT. PAT developers also provide documentation about accuracy for each country’s PAT. The results are summarized concisely in “Accuracy Results for 26 Poverty Assessment Tool Countries”

²¹ For example, the 95% confidence interval for poverty incidence is given by the 25th and 975th largest values of the 1,000 estimates. The greater the spread between these two values (i.e., the larger is the confidence interval) the less precise the tool is.

²² Recall that the validation sample is the part of the national household survey that was randomly selected to be held in “reserve” and used specifically to conduct accuracy analysis.

on the PAT website. This document provides estimates of a number of different accuracy concepts. For the purposes of this report, we maintain our focus on poverty incidence. The authors report the degree of bias, or the “poverty incidence error,” which is equal to the difference in poverty incidence as predicted by PAT from the true poverty incidence. Similar to PPI developers, PAT developers compute the poverty incidence error by comparing predicted household poverty status to true household poverty status for households in the national household survey. The accuracy document on the project website is based on in-sample results, measuring accuracy using the same households that were used in the regressions to generate the weights for the indicators. An in-sample methodology would likely result in an overstatement of the precision of the tool, because the same sample that was used to generate the tool’s weights is used to test for accuracy.²³ While PAT developers do not provide a publicly available summary of precision results, they do carry out precision analyses and provide the results upon request.²⁴ We recommend that PAT developers make both the accuracy and precision methodology and results systematically available on the PAT website.

Given space limits, this report cannot summarize the accuracy results across all countries and both tools. In general, however, **the results are quite promising**. The degree of bias in both tools appears quite low. The results of the PPI analysis of groups of households suggest that for sample sizes around 300, the bias is quite small, ranging from -2 to +2. Further, the estimates are relatively precise, with 90% confidence intervals ranging from 5 to 6 percentage points on either side of the point estimate. Similar results were obtained for the four countries for which precision results were provided by PAT developers. For example, the 95% confidence interval for bias (as measured by poverty incidence error) runs from -4.98 to 0.57 in the case of Nepal and from -1.0 to 1.6 in the case of Ethiopia. We note that, in the case of PAT, additional documentation is required—such as the year of the survey and the size of the samples upon which the precision results are based—in order to interpret these results. This documentation should be provided in detail and for each country so that users and potential users can take fully informed decisions.

In general, then, **the tools appear fairly accurate**, especially considering that a very limited number of indicators are used to predict households’ poverty status. Three caveats, however, are worth emphasizing. First, the accuracy results summarized here, and reported in detail in the tool developers’ documentation, are of course completely dependent on the quality of the household data itself. If those data were collected with significant error, then that error will be passed on to the tools. Second, the tools capture the relationship between household poverty and the selected indicators. If this relationship changes over time or varies significantly across space, then the accuracy of the tool will be reduced. Third, a tool that appears highly accurate according to statistical analysis based on national household survey data will, without a doubt, be highly

²³ PAT developers clarified, via personal communication, that they indeed test out-of-sample accuracy for PATs in most countries. We recommend that PAT developers provide greater description of the accuracy assessment methodology on the project website.

²⁴ To aid in the development of this document, PAT developers sent accuracy and precision results for the following four countries in response to our request: Ghana, Nepal, Bolivia, and Ethiopia. For these four countries, the analysis was conducted out-of-sample using 1,000 bootstrapped samples. The developers did not, however, provide any description of the methodology.

inaccurate unless the many barriers to successful MFI implementation of the tool are addressed. We return to these latter two points in the next section.

The discussion above provided one way of thinking about accuracy of PMTs. That method—which focused on poverty incidence—is appropriate when our objective is to generate a good answer to the question “What percentage of an MFI’s clients are poor?” MFIs and donors may, however, be interested in additional questions related to accuracy. For example, we might want to know the fraction of individuals that are misclassified by the tool. There are two types of misclassification: (1) “under-coverage” occurs when a truly poor household is incorrectly classified as nonpoor and (2) “leakage” occurs when a truly nonpoor household is incorrectly classified as poor.²⁵

Side-by-side comparison of PPI and PAT is also made difficult because: (1) they do not report the same accuracy tests, (2) they do not operate in all of the same countries, and, (3) they do not always use the same national data sets to construct the tool. We thus **recommend** additional statistical analysis that permits greater accuracy comparisons between PPI and PAT and continued research comparing the targeting and measurement accuracy of PMTs in microfinance. This is beyond the scope and budget of this report.

Determinants of accuracy

When considering the determinants of a tool’s accuracy, it is useful to separate ex-ante design factors versus ex-post implementation factors. The discussion until now has focused on the selection of indicators and the construction of the weights associated with each of the indicators. Our discussion has thus focused on ex-ante factors that influence tool accuracy *before it is implemented*. Ex-post factors are the factors that affect the way the tool is implemented by the MFI in the field. Here we first review the ex-ante factors discussed above and then turn to ex-post factors.

Ex-ante (design) determinants

Quality of National Household Survey

As discussed, the most important input used in designing each tool is a national household income or expenditure survey. These surveys are often long and complicated and face many challenges in implementation, including training and coordinating large survey teams, carrying out surveys in difficult-to-access and remote areas, and the entry and cleaning of massive amounts of data. For purposes of the PMTs, the most important concern is the potential that the final data set contains significant errors due to collection or data entry. If the quality the national data set is low, then the tool accuracy will also be low. For example, if expenditures are measured poorly and with a significant amount of “noise,” then the relationship between poverty status and the indicators selected in each tool will not reflect the “true” relationship in the population. Given this concern, the

²⁵ These terms originate in the targeting literature. Misclassification is a much greater concern for the targeting of social or anti-poverty programs because it implies inefficiency of government resources. See Annex 5 for a discussion on this subject.

developers should pay significant attention in evaluating the quality of these data.²⁶ This report **recommends** the tool developers to establish partnerships with institutions that collect national household data (World Bank, National Statistical Offices) and to avoid developing a PPI or PAT at country level when the underlying data set is of poor quality or too old. Tool developers could also work in collaboration with institutions developing targeting scorecards for other programs, such as the World Bank, which has developed scorecards in different countries to improve targeting efficiency of welfare programs.

Spatial differences in underlying poverty relationships

Again recall that the PMTs, via the regression methods described, seek to accurately capture the relationship between household poverty status and the indicators. In designing the tools, the developers use nationally representative data. As such, the tools capture the average relationship for the entire country. If the relationship between household poverty and the selected indicators varies significantly across regions, then this average relationship may not do a very good job of describing the relationship in a particular region. This is of potential concern to MFIs because they tend to operate in limited areas, such as a province or department, or sometimes in very remote areas. A similar concern would exist if the relationship between poverty status and the indicators is very different across rural and urban areas of a country. We note that PAT partially addresses this potential issue by including regional variables in its survey.

It is important to clarify that our concern is that the *underlying relationship* between household poverty and the tool's indicators is different in rural and urban areas. For example, if electricity is not available in rural areas but is available in urban areas, then ownership of a TV may be a very poor indicator of poverty (or lack thereof) in rural areas, although it is a very good indicator of poverty in urban areas. Now let's also assume that in this country the majority of people live in urban areas, so that the scorecard will calculate higher household expenditures or assign lower probability of being in poverty to households with a TV and lower expenditures and higher probability for households without a TV. In this case, we would be concerned that the tool would *overestimate* poverty for rural MFIs. The scorecard would incorrectly predict that even wealthy MFI clients are poor because—since they live in rural areas—they do not own TVs.

So why might the underlying relationship between indicators and poverty be different? The most common reasons are differences in access to infrastructure (electricity, water, roads, etc.) and differences in relative prices. Each of these alters the relative cost and benefits of owning different types of assets (including physical and human capital).

Note that this is a very different issue than the distribution of assets holding constant infrastructure and relative prices. For example, in many developing countries the gap between rural and urban infrastructure has significantly diminished so that the underlying relationship between poverty and the indicators will be quite similar across cities and the countryside. In this case, the tool should be equally accurate across regions. This does not imply, of course, that the poverty estimates

²⁶ The PAT developer currently screens data according to multiple factors, including the methods by which expenditure or income are collected in the questionnaire.

generated by the tool will be similar because rural MFIs may be more likely to draw from poor households than urban MFIs.

Based on our discussions with several MFIs, there is confusion on this point. MFIs operating in urban areas felt that the tools were “unfair” because they generated lower poverty estimates compared to competing MFIs in rural areas. It is important to clarify that this concern of “unfairness” is justified only if the tool’s indicators, for reasons discussed above, do not do as good a job at capturing the underlying relationship between household poverty in rural areas as the indicators in urban areas.

A **recommendation** is to conduct additional statistical analysis to evaluate the degree to which there exist spatial biases in the two tools. As discussed above, this would appear to be a potentially greater concern for PPI than PAT because PAT includes regional and rural indicators. In discussions with tool developers over this point, they indicated that designing subnational scorecards could be possible, but it is a matter of increasing costs. Moreover, subregional scorecards could add to the confusion of who is using which tool and how to interpret the scores.

Time changes in underlying poverty relationships

The concern above was that the relationship between household poverty and the tool’s indicators are different across space. This relationship may also change over time as the structure of the economy changes. The relationship may change for a number of reasons, including improvements in the quantity and quality of infrastructure, changes in relative prices, and the introduction of new goods or services. Technological change is an especially strong driver of poverty dynamics. For example, in the 1990s, ownership of a cell phone was a strong predictor of poverty (or rather of being nonpoor). Today, however, cell phones are widely available throughout urban and rural areas of developing countries and are a much weaker predictor of poverty status.²⁷

The dynamic nature of poverty (and the relationship between indicators and poverty status) has several implications for tool accuracy.

Implication 1. First, accuracy will tend to fall over time unless the tool is updated with new national household survey data. The table in Annex 3 gives the year of the national survey upon which the tools in each country are based. Both PPI and PAT in Bolivia, for example, are based on the 2007 national expenditure survey. In the Philippines, in contrast, both tools are based on the 2004 national expenditure survey. Other things equal, we would suspect that the Bolivian tools would be more accurate than their counterparts in the Philippines.

This has two implications for MFIs. First, unless the tool is updated frequently, the tool will become less accurate over time. Second, the decrease in accuracy may be systematic (i.e., over time the tool may tend to increasingly over- or underestimate poverty levels). As a result, the ability of the tool to track changes in poverty levels of clients is reduced.

²⁷ Because of the explicit goal of measuring change over time, the PPI developer seeks to avoid including indicators whose relationship with poverty is likely to change quickly over time (personal communication with PPI’s Mark Schreiner).

Just how quickly does tool accuracy fall over time (in the absence of an update)? And do older tools tend to systematically over- or underestimate poverty? The country reports for PPI provide a partial answer. In Peru, for example, the tool developer uses successive rounds of the national household sample to compare the true change in poverty rates from 2006 to 2007 to the change predicted using the 2007 scorecard applied to households from 2006 and 2007. The scorecard leads to a significant underestimate (-3.8%) compared to the true reduction of poverty rates (-6.8%). The tool developer notes that “...the main driver of this difference is probably changes in the relationship between indicators and poverty....” (Schreiner, 2009, p. 41).

This report thus **recommends** that (1) additional analysis should be conducted to evaluate the importance of this deterioration of tool accuracy over time, including how the size and direction of bias change over time, and depending on the result,²⁸ (2) additional resources should be directed to the more frequent updating of the tools.

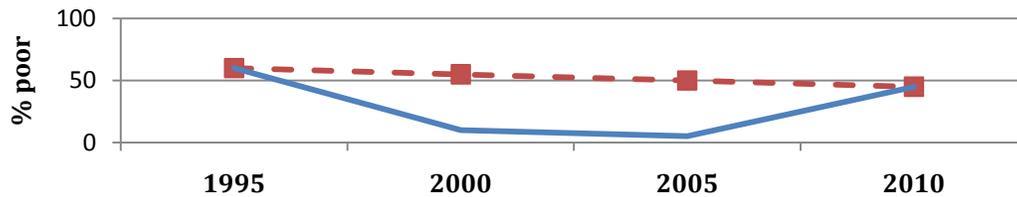
Implication 2. The transition to an updated tool raises several challenges. When a tool is updated, the developers start from scratch and construct new tools using the steps outlined above. There are several potential sources of concern when a tool is updated. First, the methodology underlying the national household survey itself may have changed over time, including sample selection, training of enumerator teams, and structure and method of implementing the survey. In this case, even if the underlying relationship between poverty and indicators does not change, the indicators selected for the new survey/scorecard and their associated weights may change. As a result, the new tool may generate a different poverty estimate compared to the old tool even though no change in poverty rates has occurred.²⁹

Figure 1 provides a hypothetical example to illustrate these issues. Assume that in 1995, ownership of a cell phone was a very strong indicator of poverty status, with only a few, relatively wealthy households having cell phones. In 1999, however, large-scale investments were made in telecommunications infrastructure with the result that cell phones and cell phone services became extremely cheap and widespread or, simply, MFIs started offering a cell-phone loan product that allowed many clients to purchase one, as our field visits confirmed. As a result, from 2000 onwards, most individuals, including the poor, owned cell phones, and cell phone ownership was no longer a good predictor of poverty. Now assume that a national household data set was collected in 1995 and again in 2010, and that a scorecard was designed in 1995 and was updated in 2010. Finally, assume an MFI adopted the tool in 1995 and applied it once every five years. The *solid line* in Figure 1 shows the poverty incidence estimates generated by the tool, while the *dotted line* shows the true poverty incidence.

²⁸ The PPI developer has carried out an analysis to evaluate the deterioration of accuracy over time due to lack of updating for 12 countries. The results are reported in the country reports available at <http://www.microfinance.com>.

²⁹ As pointed out by Mark Schreiner, this challenge is not unique to PPI or PAT, rather it is a challenge faced by other poverty measurement efforts, including the many that use purchasing power parity adjustments, which are infrequently updated.

Figure 1. Illustration of Underestimation of Poverty levels using Scorecards



The figure illustrates two specific problems. First, failure to update the tool leads to a significant bias in the estimate of poverty rates when the tool is applied in 2000 and 2005. In this hypothetical example, the true poverty incidence of the MFI’s clients has fallen slightly, from 60% in 1995 to 45% in 2010. The estimate of poverty incidence generated by the tool, however, falls significantly (from 60% to 10%) from 1995 to 2000, and then even further in 2005. The tool thus leads to a serious underestimation of poverty among clients. The reason is that the relationship between household poverty and the indicators on the scorecard has changed significantly over this period, but the tool has not been updated to reflect this change. Specifically, the tool was designed in 1995, when ownership of a cell phone was a strong indicator that a household was poor. Because of technological change or the new product of the MFI, this was no longer the case in 2000 and 2005, when cell phones were widely owned even by the poor. However, because the same tool (based on the 1995 national household data) was applied in 2000 and 2005, households that owned a cell phone in those years would be predicted as nonpoor.³⁰

The second problem is that the introduction of an updated tool may cause a biased estimate of the change in poverty of MFI clients. In Figure 1, switching to the updated survey in 2010 results in an estimated increase in poverty of 40 percentage points (from 5% to 45%) from 2005 to 2010. This change, however, is strongly biased because the 2005 estimate has a large negative bias (because it was based on the old tool) while the 2010 estimate is accurate (because it is based on the newly updated tool).

Two points bear emphasis. First, the longer a tool is used without updating with new national household data, the greater the likelihood that the poverty estimates will be biased and the larger the size of bias will be. Second, any change in methodology—including updating the tool when new national data are available and updating the tool to modify indicators or the way that survey questions are asked, or the way in which a given tool is implemented by the MFI—will likely result in a change in the estimate of poverty levels. As a result, the MFI must be cautious when interpreting estimates of changes in poverty rates.

Finally, a practical concern observed in field visits is how MFIs should operationally deal with the poverty tool update. For MFIs that have decided to roll out a scorecard across the board, with over 60,000 clients for instance, it is challenging to stop such a decision and implement the new scorecard.

³⁰ It is important to keep in mind that this is a hypothetical example and is presented only to clarify the conceptual point.

A **recommendation** is to carry out additional work to gain a better understanding of the magnitude of the biases introduced by the changing underlying dynamics of poverty. Also, additional training and guidance should be provided to MFIs on how to make the transition to an updated poverty tool.

What kinds of questions do scorecards answers?

The following types of *static*, group-based questions can be answered with a high degree of accuracy:

- **Q1:** What percentage of our *entire client base* is poor?
- **Q2:** What percentage of *new clients* is poor?
- **Q3:** What percentage of clients in *a particular branch* is poor?
- **Q4:** What percentage of *male versus female* clients is poor?

The following types of *dynamic*, group-based questions can be answered with a high degree of accuracy:³¹

- **Q5:** How does the poverty rate of our *entire client base this year* compare to the poverty rate of our *entire client base last year*?
- **Q6:** How does the poverty rate of *this year's new clients* compare to the poverty rate of *last year's new clients*?

The following type of dynamic question is harder to answer:

- **Q7:** How did the poverty rate of last year's new clients change from last year to this year?
 - Reason 1: Must apply the tool to the same sample over time.
 - Reason 2: Attrition (some of last year's new clients will not be clients this year).

The following type of individual-based questions can be answered, but with a much lower degree of accuracy:

- **Q8:** Is this individual client (or potential client) poor?
- **Q9:** Has the poverty status of this individual client changed from last year to this year?

Why are group-based poverty estimates more reliable than individual-based estimates?

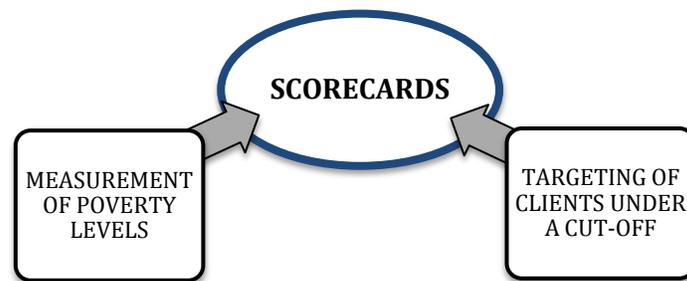
- Some truly poor are predicted to be nonpoor (undercoverage);

³¹ As discussed PPI is more likely to report changes in poverty rate over time because indicators are selected that will be more likely to change. The PPI team has provided extensive documentation about the degree of accuracy of estimates of change generated by PPI. While PAT can be used for determining change over time, its primary purpose is to measure an MFI's poverty outreach with high accuracy at a point in time.

- Some truly nonpoor are predicted to be poor (leakage);
- These errors tend to be fairly large, but they cancel each other out when we estimate the poverty rate of the entire group.

Can the scorecards be used for targeting?

The primary objective of both PPI and PAT is to **measure the poverty rates** of a clearly defined subset (or the entire population) of MFI clients at a given point in time. As such, they are primarily *ex-post* measurement tools—they are applied to individuals after (or at the moment) they become clients. In theory, each tool could also be used for *ex-ante* targeting; PPI because it generates a predicted probability that a household is poor and PAT because it generates a predicted value of household expenditures, which could then be compared to a given poverty line. The developers differ, however, in their attitudes toward the use of the tools for targeting. In their documentation, the PPI developer identifies targeting as an additional potential use by MFIs. PAT developers, in contrast, do not view their tool as an instrument for targeting, because targeting accuracy tends to be significantly lower than accuracy of poverty incidence of a group of MFI clients.



During field visits, one of the reviewed MFIs used PPI as a targeting tool; however, this usage seems to be less common. An MFI could apply the survey to potential clients and, based on the probability of being under a poverty level that is predicted by the tool, decide whether or not to consider the individual as a client. For example, assume an MFI establishes a goal of having at least 50% of new clients drawn from below the poverty line. In essence, the MFI has established an eligibility criterion, beyond the conventional criterion of creditworthiness, for the selection of new clients. The MFI could apply the survey to potential clients to separate them into those who meet the targeting criterion (i.e., those whose predicted probability of being poor exceeds a certain value determined by the MFI) and those who do not. Applicants from the poor group could then be given priority consideration in order to meet the 50% cut off.

In theory, then, the scorecard can play a similar role to the proxy means tests commonly used to target the beneficiaries of large social programs, such as conditional cash transfers.³² Indeed the statistical methodologies underlying both PPI and PAT are derived from proxy means tests.

³² Innovations for Poverty Action is currently setting a randomized control trial to measure the target efficiency of PPI in the Philippines. Other recent papers reviewing the target efficiency of different tools are Banerjee, Duflo, Chattopadhyay, and Shapiro (2007) and Alatas et al. (2010).

Any MFI considering using a scorecard for targeting should keep in mind three caveats. First, the tools are not perfect. They may lead to incorrectly identifying a truly poor household as nonpoor and vice versa. In many cases, the review team observed that MFIs keep using other screening mechanisms in parallel such as housing or food indexes, which appears to be the correct approach to avoid dependency on only one methodology, especially if that methodology (i.e., using a PMT) has a relatively low degree of targeting accuracy. Second, and most important, the tools should be considered to be only a screening mechanism; they are not a substitute for the MFI's conventional methods of evaluating the quality (for example, creditworthiness) of the household as a potential client. Finally, MFIs using the scorecards to target only very poor people should also reflect on two issues: (1) how fair it is to reject clients who are just above or below the cutoff point and (2) the financial sustainability implications of accepting only very poor clients, instead of having a portfolio of poor and less poor ones.

While targeting seems to be a minority use of the scorecards, if this use becomes more widespread, it could potentially be a differentiator between MFIs targeting all clients (more or less poor) and those that aim at a client base consisting only of clients with very low incomes.

Can the scorecards be used for impact evaluation?

The short answer is **definitely not**. During the field missions especially, confusion in this point was observed among MFIs, networks, and donors in general. Scorecards are not designed to be used as impact tools, and they should not be marketed as such.³³ In the continuum of tools that MFIs can use to target clients or to monitor and evaluate their activities, poverty scorecards are one and only one of the tools to track poverty levels and see how they change overtime.

A practical example might be useful at this point: assume an MFI implemented a scorecard on a sample of new clients in 2005 and found that 55% were poor. It then implemented the scorecard again on these same clients in 2010 and found that only 25% were poor. The MFI might be tempted to conclude that they were responsible for (i.e., they caused) this decrease in poverty. While it is of course possible that the MFI indeed caused this poverty reduction, the evidence generated by applying the scorecard on clients cannot be used to support these claims. The reason is straightforward. To attribute causal impact to the MFI, we would need to know what poverty levels of the clients *would have been if they had not become clients*. For example, it may be the case that rapid economic growth in the region was responsible for raising income levels of the poor so that many individuals would have grown out of poverty even if they had not become clients. Since the scorecard is only applied to clients, there is no way to estimate the counterfactual quantity "poverty status of current clients if they weren't clients" that is needed to say that a change in poverty was actually caused by the MFI.

It is thus important to reiterate that the primary use of scorecard is to answer questions of the form: What percentage of our clients is poor? or How has the average poverty level of our clients

³³ We do not mean to suggest that either tool is marketed as an impact evaluation tool. In fact, PPI documentation explicitly states that MFIs can use PPI to measure change but that change does not represent the causal impact of the MFI on poverty rates. The PAT Implementation Manual, website, and other documentation state this explicitly, as well.

changed over time? They are **not designed to answer questions of the form:** How did our MFI *impact* the poverty level of our clients?³⁴

³⁴ One might be tempted to apply a scorecard to a group of nonclients to estimate the counterfactual and thereby attribute causality. This would not be appropriate, however, unless the nonclients were similar to clients in all attributes related to poverty. This would be the case, for example, if households were randomly assigned to be either clients or nonclients. Households, however, are NOT randomly assigned to be MFI clients. Instead they choose (and are chosen) to become clients. Those that become clients are almost certainly systematically different from those who do not become clients. As such, they cannot be used to create the relevant counterfactual. There is a potentially important exception to this rule. Namely, if the MFI works closely with a researcher to design a research strategy, an appropriate counterfactual could be created. In this case, the scorecard could in fact be a highly valuable tool to measure the impact of MFIs on poverty. Karlan and Zinman (2010) is one example of this type of collaboration between MFIs and researchers.

Part II: User review, what do MFIs say?

4. Methodology

To gather, process, and report the user feedback contained in this report, the review team conducted a series of onsite and offsite interviews with management and staff from microfinance retail and support institutions that had used the poverty assessment tools reviewed. Tool developers were consulted to help develop a list of potential interview candidates. MFIs interviewed onsite were selected from countries where users of both tools could be visited.³⁵ This research was based on a user review methodology previously developed by a group of practitioners, academics, investors, donors, and microfinance specialists to evaluate seven social performance tools (<http://www.sptf.info/page/user-reviews-of-sp-tools>). User feedback was scored according to the methodology in Annex 9.

Each user feedback contains the following:

- User feedback on **six attributes**: Cost, Duration of Implementation, Usefulness, Robustness of Results, Ease of Use, and Overall Satisfaction
- Reported **Strengths and Weaknesses**
- Additional **User Comments**
- Review Team Feedback

³⁵ It should be noted that the Philippines and Haiti PATs are USAID-certified versions of PPI.

5. User and External Review of PPI³⁶

Progress out of Poverty Index

5.1 Summary of user comments

Box 1. Snapshot Review of Tool Attributes

<u>Cost</u>		<u>Robustness</u>	
Financial cost:	~\$1,000–\$30,000*	Perceived appropriateness of indicators:	mod/high
No. of external consultants:	0–5	Perceived appropriateness of process:	mod/high
No. of staff hired to manage:	0–1	Perceived level of confidence in results:	mod/high
<u>Duration</u>			
Training of staff and set up:	varied		
Survey time:	1–20 minutes		
Time to clean data:	varied		
<u>Usefulness to MFI</u>		<u>Ease of Use</u>	
Internal Purposes:	very useful	quite easy	
External Purposes:	useful/ very useful	<u>Overall satisfaction</u>	
		mostly satisfied	
*Figures do not include the cost of initial training to use the tool, which was subsidized. Institutions that used the tool on all clients incurred significantly higher costs than those who surveyed using samples.			

Reported Strengths and Weaknesses

Reported Strengths

- Flexibility to disaggregate data allows the user to look more deeply in analysis of the institutional position in the market, allows analysis of several poverty lines.
- Easy-to-use questionnaire is simple to implement, yields easy-to-process information that allows for quick decision making.
- Cost is low.
- Assessment allows management to know whether they are on track regarding their social mission (internal use).
- Scores are reported to social investors or disseminated at conferences or national network meetings (external use).

Reported Weaknesses

- There is limited information about the tool, how it was developed, how it should be implemented, and how the results should be used in Spanish.³⁷

³⁶ This section of the report is based on interviews conducted with MFI users of the two tools. A table of the institutions whose answers were used may be found below. While the sample is limited to seven users and not representative of all users, the authors believe their responses will be useful to an average potential user since this is the only review on the subject that the authors know of.

- The software template provided by PPI is weak, has limited report generation function, and presents compatibility issues with user MIS.
- Selection of indicators is not transparent or well explained, which reduces internal buy-in of results.

5.3 Additional User and Review Team Comments

Cost

User Comments: PPI is a relatively low-cost tool to set up and use. For those using the tool only on client samples, the major costs reported were training staff and conducting the initial sample. The only user who paid more than \$10,000 to implement PPI uses it on all clients at each cycle. Users interviewed all designed MIS software to manage PPI data internally or at low cost. Only one user hired additional staff to implement the tool on a continual basis, and only one used external surveyors to conduct field surveys. In general, user perception was that the cost of implementing PPI is minimal, with major costs being covered by the tool developer.

Reviewer Comments: While PPI was relatively cheap for the users interviewed, several variables could significantly increase the cost of implementation for other users. All users interviewed for this report received their initial PPI training for free and had staff dedicated to conducting client research. No user required significant modification of their MIS to incorporate PPI indicator management.³⁸ Also, no user paid significantly for data collection and processing; conducting the field survey of clients could be costly for other users.

Duration

User Comments: Users reported that initial training (training of trainers) took 3–5 days. Initial set up systems for implementation varied significantly depending on whether the user conducted the PPI on a sample or on all active clients. The amount of time required to collect data during client interviews was reported to be as little as one minute, up to 20 minutes. Users had systems capable of generating reports instantly once data had been entered.

Reviewer Comments: For each step of tool setup and implementation, time requirements depend heavily on who is implementing the step and how. Users interviewed received their initial PPI training from different institutions: from Grameen Foundation and from third-party institutions. The design of the curriculum, the number and competence of trainers, the selection of trainees, etc. determines how long this training (and subsequent training of staff/surveyors) lasts. The capacity of the MFI and the sample size directly affect the amount of time required for system setup. Since the PPI questionnaire is short and simple, client interviews in general are quite brief; however, some clients require more time to survey correctly (those needing more explanation, those reluctant to answer, etc.) The reviewers doubt that data for a PPI survey can be properly collected and recorded in one minute, but this is a good example of how implementation greatly affects the duration. Finally, the MFI's method of processing data

³⁷ Spanish-language materials for implementing PPI may be found at <http://www.microfinance.com>. Spanish speakers may refer to the entry for Ecuador to get a thorough explanation, which they then may apply to their own country. The users interviewed appeared to be unaware of the existence of these materials.

³⁸ A free Excel spreadsheet for data entry, information storage, and poverty rates calculation for a range of poverty lines is available at www.microfinance.com, with documentation in English as well as other languages for specific countries.

determines the timeframe for producing reports, but all of the systems reviewed (including the free Excel spreadsheet available from the tool developer) are capable of instantaneously calculating poverty probabilities.

Usefulness

User Comments: As an internal tool, users mostly rated PPI “very useful” for two primary reasons. First, it allows the MFI to gain valuable market information on its clients, both in terms of their socioeconomic characteristics as well as trends in their movement out of poverty. Armed with this information, management is able to adjust how it targets clients and adapts products and services to better meet their clients’ needs (for instance moving from individual lending to more community banks to help women specifically). Second, it gives management and staff a way to measure how well they are achieving their mission in terms of serving a particular clientele. Some MFIs use PPI together with a questionnaire to collect more information on other client aspects. The ability to combine the use of the PPI with a supplemental questionnaire appears to increase the incentives to implement the tool.

As an external tool, users rated PPI “useful” to “very useful.” The primary external utility was for attracting funding by demonstrating the MFI’s commitment and effectiveness to serving poor clients. One user claimed it could gain preferential rates based in part on statistics derived from PPI.

Reviewer Comments: Based on the MFIs visited, PPI appears to be a useful management tool allowing an MFI to analyze poverty rates among clients and make decisions to align operations with their intended mission and to improve client services. There is concern that the tool is being applied in ways that were not intended, such as impact evaluation.

One institution claimed it received preferential interest rates due to its reporting on PPI.

Robustness

User Comments: Most users rated PPI as between “moderate” and “high” for appropriateness of indicators; however, familiarity with the development of the scorecard and its underlying dataset varied greatly among users. All users disagreed with the use of one or more indicators on the scorecard. Some users were consulted in tool updates, while others were not. All users rated the appropriateness of the process of implementation as either “moderate,” “high,” or in-between. Confidence in the accuracy of the results of the tool was mostly reported as between “moderate” and “high.” Some users reported that their confidence in the accuracy of the tool differed if it was used in rural vs. urban areas.

Reviewer Comments: The review team has more concerns about the control and verification of PPI data collection and reporting at the MFI level than it does about the statistical method used, the indicators selected, or the process for calculating poverty levels. Unless PPI is implemented or verified by a qualified external agent, the accuracy of results depend on the MFI’s capacity and honesty. An MFI with low capacity could easily have significant errors in data collection and/or processing, which would be undetectable without an operational audit. An MFI with high capacity could easily manipulate a number of steps in the process to arrive at desired results, with little chance of detection.

Ease of Use

User Comments: Users praised PPI for being easy to use. The set up and training on the tool is easy. The questionnaire is short and easy to implement. Users are able to determine what question they want answered and to design their sample size to suit that purpose. English-speaking users said that there is adequate, accessible information available to clearly understand the tool and its proper implementation. Some users have direct contact with tool developers for technical assistance questions and for providing feedback on indicators. Most users reported that the tool gets easier to use over time.

Reviewer Comments: The simplicity of the tool makes it usable by MFIs of varying capacity levels. The 10-question questionnaire, if translated correctly, can be completed in a short time with clients. Most users reported no trouble in translating the questionnaire into the local languages. Results can be scored in the field and processed easily using simple, readily available software (Excel). Sufficient training on the tool is critical for appropriate implementation and understanding of results.

Overall Satisfaction

User Comments: Half of users reported that they were “mostly satisfied” with PPI, while the other half were between “mostly satisfied” and “completely satisfied.” Most users said the tool met their expectation by accurately measuring whether or not the MFI serves its target market based on client poverty levels.

Reviewer Comments: User satisfaction with PPI was high, generally because it is cheap, is easy to use, and accurately measures client poverty levels.

Table 1. Users Interviewed for Review

MFI	No. of Borrowers	Loan Portfolio (US\$)	Regulatory Status	Used Other Poverty Assessment Tools?
ASKI (Philippines)	41,303	8,622,392	Unregulated NGO	Yes (means testing)
FONDESURCO (Peru)	4,939	7,765,588	Unregulated NGO	No
NWTF (Philippines)	84,958	9,400,000	Unregulated NGO	Yes (cash flow analysis/means testing)
PRISMA (Peru)	21,455	10,461,655	Unregulated NGO	Yes (means testing)

Source: Mix Market (<http://www.mixmarket.org>, accessed April 2010).

5.4 Review Team Feedback

PPI is a simple, relatively low-cost poverty assessment tool that is appropriate for users who want to estimate the likelihood that a client falls below a poverty line (or multiple poverty lines), or to estimate the percentage of clients in the portfolio that fall below the line (or lines). When used correctly, PPI is quite useful for assessing whether the MFI is fulfilling its mission by reaching clients within certain poverty levels and for gaining valuable market information on clients. The tool cannot be used in countries where no PPI scorecard has been developed.

The accuracy of PPI results depend heavily on the capacity and honesty of each user. Among users interviewed there are few controls in place to detect significant errors, either intentional or unintentional. Users are currently conducting little or no PPI software auditing—the integrity of the data relies on the sophistication and robustness of the MIS, which varies significantly among MFIs. There is currently little systematic internal or external verification or validation of tool implementation and reported results.

The quality of PPI training received was inconsistent among interviewed MFIs: some received training directly from Grameen, others from a third party. When training quality and completeness were high, users had a better understanding of how the tool functions and its proper application.

PPI results in the same country are not necessarily comparable for several reasons. MFIs may be using different versions of PPI or they may be using the same version of the tool, but on different sample groups (new clients, all clients, specific geographic regions of the country, etc.).

Keeping PPI indicators up to date presents a real challenge for tool developers. Some indicators can quickly become inappropriate, particularly those based on ownership of an asset (such as a mobile phone, a television set, etc.) as explained in Part I of this report. The process of rolling out new scorecards is time consuming and requires both field-testing by users and explanation of changes. Failure to do both can result in inaccurate results or low user acceptance. It is important that users understand how old indicators are and whether or not they have been field tested before using.

It is important to clearly communicate to users (MFIs and donors) what the tool can and cannot do, as well as how to verify the accuracy of results. Some users disaggregate the data and interpret results in novel ways. These uses were not necessarily the intent of the developers and have not been assessed for validity.

6. User and External Review of PAT

Poverty Assessment Tool

6.1 Summary of user comments:

Box 1. Snapshot Review of Tool Attributes

<u>Cost</u>		<u>Robustness</u>	
Financial cost:	\$1,500–\$10,000*	Perceived appropriateness of indicators:	mixed
No. of external consultants:	0–4	Perceived appropriateness of process:	moderate–high
No. of staff hired to manage:	0–2	Perceived level of confidence in results:	mixed
<u>Duration</u>		<u>Ease of Use</u>	
Training of staff and set up:	1 wk–4 mo		somewhat easy
Survey time:	10–40 min	<u>Overall satisfaction</u>	
Time to clean data:	< 1 day		mostly satisfied
<u>Usefulness to MFI</u>			
Internal purposes:	mixed		
External purposes:	mixed		

*Figures do not include the cost of an initial training of trainers attended by a representative from each organization, which was paid for by USAID.

6.2 Reported Strengths and Weaknesses

Reported Strengths

- PAT is quite inexpensive to use.
- The manual and instructions for implementation are clear and complete, which makes the tool easy to implement.
- The help desk is responsive and capable of resolving issues not covered in the manual or those that arise during implementation.
- Scores can be used to report to social investors and donors or can be disseminated at conferences, national network meetings (external use of scores), etc.

Reported Weaknesses

- The tool yields only one result, which limits its effectiveness as a management tool.
- It is unclear why the specific indicators are considered valid for measuring the incidence of poverty.
- While the tool is a USAID requirement, no budget was given for implementation, which means that projects have to take the money from other budget line items.

6.3 Additional User and Review Team Comments

Cost

User Comments: Users considered the tool inexpensive relative to other tools intended to measure

socioeconomic data on clients. Most users were required by USAID to implement the tool, but none was given a budget to do so. This meant that they had to take money budgeted for other activities to pay for PAT.

Reviewer Comments: Considering that USAID paid for the initial training for all users, the main drivers of financial costs for implementing PAT are the cost of using external surveyors (if used) and the cost of conducting client visits. For apexes that have to visit clients of various institutions with a wide geographic spread, this can be quite expensive.

Duration

User Comments: Most users said PAT was quick to set up and took less than two months. One user had experience in using the tool before, so this user was able to set up implementation in one week. The time required to survey one client is usually between 10 and 20 minutes; however, some clients take longer. Once the information has been properly collected, data entry, verification, and reporting are quickly done; results come “at the push of a button.”

Reviewer Comments: All users implemented PAT on samples of fewer than 500 clients. Apexes implement PAT on the clients of partner institutions, which requires coordination to select and visit client samples. If institutional capacity is high among partners, this may be done rapidly. If capacity is low, this is time consuming. One user trained partner institutions to conduct client interviews themselves, which reduced the time required for implementation.

Usefulness

User Comments: All users reported that PAT’s usefulness is limited by the fact that it produces one result: a single percentage against one poverty line. Half of the users found the tool “useful” for internal purposes, reporting that it helped them to gain knowledge of their clients and allowed them to save money compared to other assessment tools. One user found the tool “very useful,” while another rated it as “not very useful” internally. Users found the tool less useful for external purposes. Most report PAT results to USAID only, which provides little or no feedback.

Reviewer Comments: The closed software and limited reporting function lowers the usefulness of the tool. While it is possible to conduct additional analysis, either in Epi or by exporting data to other programs, all users interviewed said that they were unable to disaggregate data or measure against different poverty lines. Some users and networks are modifying the tool to use in their own way, and half of those interviewed added a rider to PAT to collect and analyze information that they consider more useful. The next version of the tool will offer the ability to measure against three poverty lines and will allow users to conduct additional analysis and desegregation of the data, which will address users’ biggest complaints.³⁹

Robustness

User Comments: All users had doubts on the appropriateness of some of the indicators used. The opinion on the appropriateness of the process of collecting and processing the data was either “moderate,” “high,” or in between. The confidence in the results of the tool varied greatly (“low” to

³⁹ Fourteen of the current PATs already include more than one poverty line, and the remaining existing tools will be updated to add at least one line. The PAT development team reports that future tools will have the ability to measure against two to five poverty lines and allow users to conduct additional analysis and desegregation of the data. If this happens, it will address perhaps the biggest complaints about the tool.

“high”), primarily based on the acceptance of the indicators used.

Reviewer Comments: Most users did not have a good understanding of how the tool developers arrived at the indicators selected. Those who accepted the validity of the selection “on faith” had a higher opinion of the appropriateness of the indicators, as well as of the final output.

There was little verification of the survey methodology applied by the four institutions interviewed. Additionally, there was little indication that USAID does anything with the data besides “check the box” that a requirement has been fulfilled. If that is the case, it is doubtful that resources will be invested to ensure that information is correctly collected and processed.

Ease of use

User Comments: Most users said the tool is “somewhat easy,” while one rated it as “quite easy.” Users said that PAT training is good and sufficient for proper implementation of the tool. The user manual and implementation instructions are clear and complete and facilitate preparation and implementation of the tool. All users noted that the help desk is quite responsive and capable of resolving issues not clearly explained through the manual.

Reviewer Comments: The instruction and support for implementing PAT seem to be quite complete, and users can be trained through a seminar, through online courses, or by reading the manual. However, materials are primarily available only in English. (Instructions are in Spanish for some countries, and there is a Spanish version of the original implementation manual, although the recently updated version of the manual has not yet been translated.)

Overall satisfaction

User Comments: Most users said they were “mostly satisfied” with PAT, though most also said that if it were not a USAID requirement, they would not continue using the tool. Users expected that using PAT would help them to learn more about the socioeconomic level of their clients, and most believed their expectations were met. Most users cited the tool’s limited functionality as a source of dissatisfaction.

Reviewer Comments: As a requirement, PAT is not overly burdensome for most users in terms of cost, time, and human resources. Overall satisfaction appears to mirror the users’ opinion of the appropriateness of indicators used in the tool.

Review Facts

Table 1. Users Interviewed for Review

	No. of borrowers	Loan portfolio (US\$)	Regulatory status	Used other Poverty Assessment Tools?
MFI				
DAI (Haiti)*	279,576	NA	APEX	Yes
LEAD (Uganda)	NA	NA	APEX	Yes

Microenterprise Access to Banking Services (Philippines)*	700,000	62,108,604	APEX	Yes (means testing)
PRISMA (Peru)	21,455	10,461,655	Unregulated NGO	Yes (household income/expend analysis)

Source: Mix Market (<http://www.mixmarket.org>, accessed April 2010), except where indicated by an asterisk. For apexes, the number of borrowers and loan portfolio reported are for partner MFIs.

6.4 Review Team Feedback

PAT is a simple, relatively low-cost poverty assessment tool that is appropriate for USAID partners that are required to report poverty outreach indicators to comply with U.S. Congressional requirements.

PAT developers have focused on making the tool accessible for users and on controlling aspects of implementation and use: a variety of resources are readily available to help users implement the tool.

- The manual and instructions for implementing PAT are clear and complete.
- User support is available through multiple channels: an online training course, an offline USB-base training course, a live training course, an implementation manual, and a responsive Help Desk.
- Data entry template and software are based on software that is both free and specifically designed for storing, managing, and analyzing survey data. No new IT system and, in most cases, very little IT support is necessary.
- PAT software is closed, so that calculation of results is consistent and controlled by tool designers.

PAT was designed to calculate poverty at an aggregate level against a single poverty line. The next version of PAT will have the ability to measure against three poverty lines and allow users to conduct additional analysis and desegregation of the data. If successfully implemented, it will address perhaps the biggest limitations of the tool: that its usefulness for analysis and report generation is limited.

Users seem to lack full understanding of why the indicators selected are valid. As a result, they either have faith that they are appropriate or lack confidence that the tool yields valid results. As with PPI, keeping PAT indicators up to date presents a real challenge for the developers. Some indicators can quickly become inappropriate, particularly those based on ownership of an asset (such as a mobile phone, a television set, etc.) as explained in Part I of this report.

It is unclear how USAID uses the results, other than to fulfill a congressional mandate, as the users interviewed had received no feedback on their reports. This, coupled with the fact that there is little verification of process or results, could lead to poor implementation.

7. Summary table

Comparative Review Matrix of Scorecards

Tool attributes	PPI	PAT
1. Costs		
Financial cost of implementing tool	\$1,000–\$30,000	\$1,500–\$10,000
No. of external consultants	0–5	0–4
No. of staff hired to manage	0	0–2
2. Duration		
Training of staff and set up	Varied	1 week to 4 months
Survey time with client	1–20 minutes	10–40 minutes
Time to clean data	Varied	< 1 day
3. Usefulness to MFI		
Internal utility	Very Useful	Mixed
External utility	Useful/Very Useful	Mixed
4. Robustness		
Perceived appropriateness of indicators	Moderate/High	Mixed
Perceived appropriateness of process	Moderate/High	Moderate/High
Perceived level of confidence in results	Moderate/High	Mixed
5. Ease of use		
	Quite Easy	Somewhat Easy
6. Overall satisfaction		
	Mostly Satisfied	Mostly Satisfied
Note: Figures do not include the cost of initial training to use the tool, which was subsidized. Scoring based on reviews of seven MFIs.		

8. Recommendations

We conclude by offering the following recommendations. While PPI and PAT have been used by a small number of MFIs (according to MIX) and the area of poverty measurement and targeting in microfinance is still under development, there is enormous potential for roll out the use of tools as more donors are looking for poverty-level data and an increasing number of social investors in microfinance need to show (social) results and monitor their financial and social performance. The tools appear to be cost-effective alternatives for responsible financial services' providers that have the express, stated intention to serve poor clients and want to measure whether they fulfill that goal.

Recommendations for tool improvement

- Field testing is key to increasing the accuracy and uptake of scorecards. The reviews revealed that tool developers do not have enough time or resources to field test the tools. For example, there is a PPI for Malawi, but it has not been field tested. It is critical to have a proper, supervised field test of each scorecard before rolling it out. The tool developer and tool sponsor strategy to consult with donors, MFIs networks, and investors before rolling out the tools is unclear and ad-hoc.
- Opportunities for increased joint tool development among PPI, PAT, and any other scorecards developed country-wide (i.e., World Bank scorecard for Pakistan Benazir Income Support program) should be explored. Coordination of scorecard development by these three entities (USAID, Grameen, and the World Bank) will avoid overlapping and will make the best use of resources. These resources could be used to improve the way scorecards are implemented, which is the main issue of concern highlighted in this report. This report recommends organizing a workshop with all the main actors involved in scorecard development at the country level to discuss joint tool development and communication. The objective of the workshop would be to discuss the feasibility of having one common set of indicators for PPI and PAT per country.
- The reviews evidenced very low external validation systems in place and room to improve the monitoring and oversight of poverty scorecard data collection. Social raters⁴⁰ need to sharpen their social rating product and harmonize validation methodologies. Temporary systems of external validation⁴¹ could be helpful initially but the final goal should be to promote a market of improved and simplified social ratings. Social investors have a key role in guiding raters in this process.
- Additional work should be conducted to evaluate how quickly the tool's accuracy decreases over time when it is not updated with new national survey data. If accuracy erodes rapidly, then developers should consider suspending the tool or at least warning MFI's and donors about the strong potential for the introduction of bias. Many of the tools currently being used are based on national surveys over five years old and some as old as 12 years, raising doubts about their overall validity. Overall, increased research on targeting and measurement efficiency of PPI, PAT, or other available scorecards is a priority.

⁴⁰ There are four specific microfinance social rating companies: M-Cril, Planet Rating, Microfinanza, and Microrate. Moody's is moving into this area and plans to launch a rating methodology by the end of 2010.

⁴¹ As part of its Bankers without Borders program, Grameen Foundation plans to use pro bono consultants to verify correct PPI implementation.

- Based on conversations with several MFI's, it is clear that the implementation of the tools by different MFIs is extremely heterogeneous. While heterogeneity need not be negative (since objectives across MFIs may be different), it would appear that the quality of implementation is highly variable. We thus recommend that the tool developers, perhaps in coordination with donors, increase the amount of training directly to MFIs and ensure that any MFI that implements the tool (at least for external reporting purposes) meets minimum quality standards.
- The introduction of an updated tool to reflect newly available national data sets creates an additional challenge. Specifically, the MFI will not be able to distinguish what part of the change in poverty incidence from the previous year was due to true changes in poverty versus that due to the introduction of the new tool. One possible solution is to implement both the old and the new tool on separate random samples and compare poverty incidences.

Recommendations on data management

- There were several data interpretation issues. The review team observed incorrect and unrelated-to-the-scorecard information analysis done by MFIs. There is a risk of over-interpretation of information on scorecard results. During interviews MFI management explained that household asset information (knowing whether the client has a TV or not, or whether the kids are at school or not) allows them to do better product development. More research on how MFIs develop better products thanks to scorecard information or whether this is possible at all should be done. In general, improved communication on what the tool does or does not do is required.
- MFIs that are not using scorecards with all clients but with a subset of their clientele need more guidance on how to design statistically strong samples. An instructional guide in this area would help. During field reviews it was observed that MFIs do not know how to adjust their sample designs or design them in a way that allows for longitudinal analysis. While this skill depends very much on the local resources available within the MFI, tool developers should strengthen their guidance on this point to MFIs. A key next step is tool developer collaboration in creating subroutines for data analysis.
- In regard to MIS systems, MFIs implementing the scorecards face the challenge of managing the collected data. This is an issue especially for those who do not use a sampling approach but implement the scorecard with all clients. Diverting IT resources to a parallel system of social data collection might not be a good solution, and donors supporting specific SPM data management systems should evaluate whether the MFI needs a holistic support on data collection and management and not only on social data.
- In regard to the publication of scorecard data: MIX has been collecting social performance data of MFIs since 2008.⁴² At the end of 2010 more than 300 MFIs will have reported on a set of qualitative and quantitative indicators related to the social mission of a given MFI. PPI and PAT scores are part of that reporting framework. The review team observed during field missions

⁴² <http://www.mixmarket.org/social-performance-data>

that some MFIs that piloted PPI and obtained low scores are resistant to publish results and to continue tracking poverty levels. This is a trend to monitor closely, especially by donors or investors subsidizing tool uptake. Compulsory publication of scores should be part of the conditionality of financial support for tool roll out.

Recommendations for speeding up the roll out and use of scorecards

Investors

- A communication campaign targeting investors on the available tools to track how their investees measure client poverty would be useful. Tool developers could start offering training to social investors, for instance.
- MFIs' incentives to implement scorecards may increase if investors start asking for and using scorecard information (during due diligence, for instance). An important first step would be for social investors to track whether the investee measures poverty levels, how and which tool they use to target clients. Up to now only a handful of social investors sponsor or use PPI or PAT. This report recommends reviewing the tools and information that (social) investors ask their investees to report in the area of poverty measurement.
- The reviews emphasize the interpretation problems of PPI and PAT data by both MFIs and investors. Investors interested in poverty measurement among their investees could promote tool uptake. This report encourages some of the investors to fund and financially support the roll out of scorecards among their investees.

Donors

- One of the findings of this review is that there is low usage of absolute poverty tools in microfinance. While it is impossible to know how many MFIs are currently using PPI and PAT, fewer than 30 reported having used or piloted the PPI scorecard to MIX. Forty MFIs and enterprise development providers have implemented PAT for USAID reporting. To increase uptake, the use of PPI and PAT should be expanded beyond Grameen or USAID affiliates by creating partnerships with other networks and promoting the tools among other social businesses.
- MFIs interviewed show very different levels of human and financial resources to implement scorecards. In general, while some MFIs are able to implement the poverty scorecard effectively, more support is needed to help them interpret and use the scorecard data internally.
- One of the most pressing issues for poverty scorecard development is the low quality or inaccessibility to national household sets. Donors interested in funding PPI or PAT could help tool developers access the latest available national household surveys by leveraging their country or government relations.
- A donor committee could be established to coordinate and pool funding efforts on MFI poverty outreach. While funding for developing scorecards at the country level in countries with high levels of microfinance penetration is important, the low uptake at the institutional level

(according to MIX) calls for a collective effort to increase use among interested MFIs. The facility could fund some of the different activities:

- Implementation of PPI/PATs (to monitor poverty levels) in Grameen/USAID partners and non partners. One channel to funnel this support could be microfinance networks.
- MFIs' MIS improvement to facilitate PPI/PAT data management and analysis. MIS support should focus on the collection of financial and social data without creating parallel MIS for scorecard data collection.
- Training of MFI staff in monitoring and evaluation techniques— or sponsoring of their participation in available trainings.
- Establishment of a pool of certified consultants that can implement PPI/PAT across countries at the MFI level.
- Funding of updates of questionnaires to account for changes in economic indicators.

9. Annexes

ANNEX 1: Poverty Assessment Tool for Peru (2007)

Client Assessment Survey - Peru

Survey Number _____

Interviewer: Fill out the information below before the survey begins. Do not ask the respondent for this information.

Date of Interview

Interviewer (code)

Branch (code)

Location/Region

1=Lima Metropolitan; 2=Urban Coast; 3=Rural Coast; 4=Urban Highlands
5=Rural Highlands; 6=Urban Lowlands; 7=Rural Lowlands

Client Location: Urban=0; Rural=1

Time in Program: Months

Client or ID #

Quality Control Checks	
<i>Field Supervisor</i>	
Date _____	Initials _____
<i>Data Processing Coordinator</i>	
Date _____	Initials _____
<i>Data Processor</i>	
Date _____	Initials _____

Interviewer: Introduce the survey to the respondent. Say: "Hello. My name is _____. I work for the organization _____. We are trying to learn a little bit more about the clients we work with, and so I have a few questions I would like to ask today. It should only take us about 20 minutes, and the answers you provide will be put together with answers from other households. All of your answers are completely confidential and your name will not be given with your answers. Are you willing to take some time to answer these questions today?" After he/she agrees, proceed with the dialogue below.

Interviewer: "I would like to ask you some questions about the people in your household. Let me tell you a little bit about what we mean by 'household.' For our purposes today, members of a household are those that live together and eat from the "same pot." Each person contributes to and benefits from the household. It should include anyone who has lived in your house for 6 of the last 12 months, but it does not include anyone who lives here but eats separately. Do you have any questions about that?" Answer any questions the respondent has before proceeding.

Interviewer: Ask the respondent for the information in Columns A-E, as it pertains to each person in his/her household. Write the information down in the chart as he/she relays it to you. Say to the respondent: "Now I would like you to identify each person in your household and answer some basic questions about each person."

A. Household Member	B. Sex	C. Relation to household head	D. Age (complete years)	E. Education (highest class passed)
1.	0=Female; 1=Male	1=head; 2=spouse; 3=child; 4=parent; 5=grandchild; 6=grandparent; 7=other		Enter the following codes: Never attended or pre-school only=0; Primary attended=1; Primary completed=2; Secondary attended=3; Secondary completed=4; Superior no universitario (attended or completed)=5;
1) Respondent				
2)				
3)				
4)				
5)				
6)				
7)				
8)				
9)				
10)				
11)				
12)				
13)				
14)				
15)				

Interviewer: Skip the following section and return to fill in the answers after the interview. Do not ask the respondent these questions; fill in the answers from the information in the table above.

- Survey Number
2. Number of people living in household (record total number of names from Column A above):
3. Age of Household Head (record age of person from Column D, who is household head in Column C):
4. Number of other household members (not including head) who have no education (number of persons with education code of 0 in Column E above):

Interviewer: "Now I would like to ask you a few questions about your house and some items you may own."

5. How many rooms are there in the house where you and your family live? Number
(Interviewer: Include detached rooms in same compound if same household. Exclude bathrooms, toilets, kitchen and basement)

6. Do you have a telephone inside your house? (fixed landline, excluding community telephones)
 0=no, 1=yes

Interviewer: For questions with multiple choice answers, do not read the answers. Ask respondent the question and match the answer to the option on the survey most similar. If respondent's answer is unclear, probe until you find an adequate answer.

7. What type of roofing material is used in your house?
 1. Leaves 3. Straw 5. Tiles 7. Brick/cement
 2. Jute stick 4. Bamboo/wood 6. Cl sheet (corrugated tin)

8. What type of exterior walls does your house have?
 1. No walls (jungle) 2. Wood 4. Sticks with dirt 6. Bricks
 1. Woven jute stick 3. Adobe (lime bricks) 5. Stones with dirt

9. What is your primary source of drinking water?
 1. Dam, pond or river 5. Public well
 2. Rainwater collected at/near house 6. Untreated piped (river) water
 3. Water is trucked in 7. Treated piped water in residence yard (shared)
 4. Public borehole (open), Spring 8. Treated piped water in residence (own)

10. What kind of lock does the main entrance door of the house have?
 1. No lock 3. Key lock or simple padlock
 2. Wood or metal bar to close from inside only 4. Security key lock/metal frame with padlock

11. How many cars does your household currently own? Number

12. How many color TVs does your household currently own? Number

13. What is the total number of metal pots owned by your household? Number

Interviewer: The following two questions ask for both the number of an item owned and the total value. Put the number in the corresponding blank and the soles value in the corresponding box. If a respondent does not own an item, mark "0" in the soles value box.

14. How many tractors and trucks does your household own? _____
 What is their total resale value at the current market price? Soles

15. How many electric food processors does your household own? _____
 What is their total resale value at the current market price? Soles

Interviewer: Please make sure that the setting of the interview ensures confidentiality before beginning this section. Say: "I know that the following questions are sensitive. I assure you that the answers will not be shared with anyone else."

16. During the last 3 months, have you or anyone in your household received in-kind services from food aid programs (e.g., Vaso de leche, comedores, Desayuno Escolar, etc.)?
 0=no, 1=yes

17. How many members of your household belong to a water/waste group? Survey Number
(water and wastewater groups, such as Junta de Administración de Agua y Saneamiento)

18. Do you, your spouse or anyone else in your household have a withdrawable savings account, checking account, a fixed-term deposit account, or any other type of savings account?
(If yes, then enter a "0" for the next question (#19) and end the interview) 0=no, 1=yes

19. Why do neither you nor anyone else in your family have a savings or checking account?
0. Have an account 3. Institutions are not safe 6. There are no institutions available
1. Too little income (cannot save) 4. Interest rates are too low
2. No habit of savings 5. Do not know where or how to get an account

Interviewer: Look over the survey to see if you have missed any questions. If you have, please ask those questions of the respondent. If not, it is the end of the interview. Remember to thank the respondent for his/her time in helping you answer these questions!

Now return to the questions in the box below the roster and fill in the answers.

USAID Poverty Assessment Tool, Peru, Version 1.3

Source: The Peru Poverty Assessment Tool is available at <http://www.povertytools.org/tools.html>

ANNEX 2. Progress out of Poverty Index for Peru (2009)



Figure 1: A simple poverty scorecard for Peru

Entity	Name	ID	Date (DD/MM/YY)
Member:	_____	_____	Joined: _____
Loan officer:	_____	_____	Today: _____
Branch:	_____	_____	Household size: _____

Indicator	Value	Points	Score
1. How many household members are 17-years-old or younger?	A. Four or more	0	
	B. Three	6	
	C. Two	10	
	D. One	18	
	E. None	28	
2. What is the highest educational level that the female head/spouse completed?	A. None, pre-school, or kindergarten	0	
	B. Grade school (incomplete)	6	
	C. Grade school (complete)	9	
	D. High school (incomplete)	11	
	E. Non-university superior (incomplete) or no female head	12	
	F. Non-university superior (complete) or higher	19	
3. What is the main material of the floors?	A. Earth, other, or no residence	0	
	B. Wood planks	1	
	C. Cement	3	
	D. Parquet, polished wood, linoleum, vinyl, tile, or similar	17	
4. What is the main material of the exterior walls?	A. Adobe, mud, or matting	0	
	B. Wattle and daub, wood, matting, brick or cement blocks, stone blocks with lime or cement, other, or no residence	3	
5. Excluding bathrooms, kitchen, hallways, and garage, how many rooms does the residence have?	A. One	0	
	B. Two	1	
	C. Three or four	5	
	D. Five	6	
	E. Six or more	10	
6. What is the fuel used most often in the household for cooking?	A. Firewood, charcoal, does not cook, or other	0	
	B. Kerosene, gas, or electricity	5	
7. Does the household have a refrigerator/freezer?	A. No	0	
	B. Yes	5	
8. How many color televisions does the household have?	A. None	0	
	B. One	3	
	C. Two or more	8	
9. Does the household have a blender?	A. No	0	
	B. Yes	3	
10. Does the household have an iron?	A. No	0	
	B. Yes	2	

Total score: _____

Microfinance Risk Management, L.L.C., <http://www.microfinance.com>

60

For more information visit www.progressoutofpoverty.org

Source: Schreiner (2009), p. 60.

ANNEX 3. Countries with PAT and PPI

Year of National Household Data upon Which Tool Is Based

Country	PAT	PPI
Albania	2002	NA*
Azerbaijan	2002	NA
Bangladesh	2004/2005	2004/2005
Bolivia	2005	2007
Bosnia & H.	2004	NA
Brazil	NA	2008
Cambodia	2004	2004
Colombia	2008	NA
East Timor	2001	NA
Ecuador	2006	2006
Egypt	NA	2004
El Salvador	NA	2008
Ethiopia	2004	2004
Ghana	2006	2006
Guatemala	2000	2006
Haiti	2001	2001
Honduras	NA	2007
India	1998	2006
Indonesia	2000	2007
Jamaica	2000	NA
Kazakhstan	2004	NA
Kenya	NA	1997
Kosovo	2000	NA
Madagascar	2004	NA

Malawi	2005	2005
Mali	NA	2001
Mexico	2008	2008
Morocco	NA	1999
Nepal	2004	2004
Nicaragua	NA	2009
Nigeria	NA	2003
Pakistan	NA	2006
Palestine	NA	2005
Paraguay	2002	NA
Peru	2007	2007
Philippines	2004	2004
Romania	NA	2007
Russia	NA	2007
Senegal	NA	2005
Serbia	2007	NA
South Africa	NA	2006
Tajikistan	1999	NA
Uganda	2004	NA
Vietnam	1998	2006
West Bank	2007	NA
Yemen	NA	2006

*Indicates the tool is not available in the country.

ANNEX 4: PPI Lookup Table for Peru

Figure 4 (National poverty line): Estimated poverty likelihoods associated with scores

If a household's score is then the likelihood (%) of being below the poverty line is:
0-4	96.3
5-9	90.4
10-14	91.0
15-19	80.7
20-24	75.4
25-29	62.1
30-34	51.4
35-39	38.1
40-44	27.2
45-49	14.9
50-54	11.3
55-59	6.4
60-64	3.3
65-69	0.2
70-74	0.0
75-79	1.5
80-84	0.0
85-89	0.0
90-94	0.0
95-100	0.0

Surveyed cases weighted to represent households in Peru.

Based on the 2007 ENAHO.

Source: Schreiner (2009), pp. 60 and 73.

ANNEX 5: Discussion on Leakage and Undercoverage

Case 1: Perfect Measurement

		True Status	
		Poor	Non Poor
Predicted Status	Poor	10	0
	Non Poor	0	10

Case 2: Imperfect Measurement

		True Status	
		Poor	Non Poor
Predicted Status	Poor	0	10
	Non Poor	10	0

In both cases, the tool was applied to 20 households. In each case, the true poverty incidence is 50%. In Case 1, all 20 households' poverty status is *correctly* predicted. In Case 2, all 20 households' poverty status is *incorrectly* predicted. Should we be concerned with the large leakage and undercoverage of Case 2? Not if our only goal in applying the tool is to estimate poverty incidence! Although many more individual households are incorrectly classified in Case 2, in both cases the tool generates exactly the same estimate of poverty incidence. There may be circumstances, however, under which an MFI or a donor would want to go beyond poverty incidence. An MFI that wants to use the tool for targeting, for instance, would not be indifferent to the two cases presented. Instead, it would like to know the percentage of households that were correctly identified as poor and nonpoor. Both tool developers also report a number of additional accuracy measures, including leakage, undercoverage, and a relatively new criterion developed by PAT developers called the Balanced Poverty Accuracy Criterion, which incorporates targeting errors and poverty incidence. Given space limitations, this report does not analyze the tools based on these alternative notions of accuracy. Reader are referred to the PAT developers' document, "Note on Assessment and Improvement of Tool Accuracy," available on the PAT website.

ANNEX 6. Accuracy Results for 2005 PPI for Bangladesh

Figure 8 (\$1.25/Day 2005 PPP poverty line): Bootstrapped differences between estimated and true household poverty likelihoods with confidence intervals in a large sample (n = 16,384), 2005 scorecard applied to the 2005 validation sample

Score	Difference between estimate and true value (2005)			
	Diff.	Confidence interval (+/- percentage points)		
		90-percent	95-percent	99-percent
0-4	+0.0	0.0	0.0	0.0
5-9	-2.0	2.3	2.9	3.6
→ 10-14	-5.9	3.5	3.6	3.7
15-19	+2.9	1.4	1.7	2.3
20-24	+0.9	1.9	2.2	3.1
25-29	-4.4	3.1	3.2	3.5
30-34	-3.7	2.9	3.1	3.6
35-39	-3.7	2.9	3.0	3.3
40-44	-2.4	2.2	2.5	3.3
45-49	-2.0	2.1	2.4	3.1
50-54	-1.1	2.2	2.6	3.4
55-59	+1.3	1.6	1.9	2.7
60-64	+1.1	1.7	2.1	2.6
65-69	-1.8	1.6	1.8	2.2
70-74	-0.8	0.8	0.9	1.2
75-79	+1.0	0.5	0.6	0.8
80-84	-8.8	6.6	7.0	8.3
85-89	+0.0	0.0	0.0	0.0
90-94	+0.0	0.0	0.0	0.0
95-100	+0.0	0.0	0.0	0.0

Figure 10 (\$1.25/Day 2005 PPP poverty line): Differences and precision of differences for bootstrapped estimates of poverty rates for groups of households at a point in time, by sample size, 2005 scorecard applied to the 2005 validation sample

Sample size (n)	Difference between estimate and true value			
	Diff.	Confidence interval (+/- percentage points)		
		90-percent	95-percent	99-percent
2	-0.5	49.9	58.5	73.5
4	-1.8	32.9	39.3	50.6
8	-1.6	24.0	27.5	37.7
16	-1.4	17.6	20.6	25.7
32	-1.4	11.7	13.7	18.8
64	-1.5	8.6	9.7	13.0
128	-1.6	6.0	6.8	8.3
→ 256	-1.6	4.2	4.9	6.4
512	-1.7	2.9	3.5	4.8
1,024	-1.7	2.1	2.4	3.3
2,048	-1.7	1.4	1.7	2.3
4,096	-1.7	1.1	1.3	1.6
8,192	-1.7	0.7	0.9	1.1
16,384	-1.7	0.5	0.6	0.8

Source: Chen and Schreiner, 2009, p. 112.

ANNEX 7. Details of Construction of PAT and PPI

This section provides greater details on how PPI and PAT are constructed. It is based on the documentation of the two tools and responses of the tool developers to a developers' questionnaire.⁴³

Tool construction: PPI

- **Indicator Selection.** The following steps are used to select the final set of indicators in PPI.
 - The developers first identify the highest quality and most recent national survey data that contain information on expenditures or income.
 - PPI developers seek to identify 10 indicators that will appear on their “poverty scorecard,” the survey form that will ultimately be applied to client households.
 - The first step is to eliminate those indicators that are not easily verifiable. This exclusion is based on common sense and the developers’ experience in conducting poverty surveys. The remaining pool of indicators typically includes about 100 candidates.
 - Randomly separate the full sample from the national household survey into three parts: Logit sample, calibration sample, accuracy assessment sample.
 - Compute the poverty status (1=poor, 0=Nonpoor) for all households according to the national poverty line.
 - The next step is to build a one-indicator “scorecard” for each of these candidate indicators. This involves the following steps:
 - Using the Logit sample, run a logit regression of poverty status against the (single) indicator.
 - Repeat this regression for each of the candidate indicators.
 - The next step is to select one of these one-indicator scorecards. The selection criteria balance accuracy (which will be described in greater detail in the next section) against the practicality issues discussed in the first section. The one-indicator is selected as follows:
 - Convert the logit parameter estimates into integer values—or points—such that when they are summed up, the maximum score is 100 and the minimum score is 0.
 - Use these points to compute the poverty score for all of the households in the calibration sample.
 - Use the calibration sample to generate poverty probabilities associated with each possible score (or range of scores). For example, the probability of being poor for a household with a score between 20 and 22 is equal to the fraction of households in the calibration sample with a score between 20 and 22 that was observed to be poor.
 - The next step is to construct a series of two-indicator scorecards, each one adding a different indicator to the indicator selected in the previous step. After a similar evaluation process balancing improvements in accuracy versus practicality, the best two-indicator scorecard is selected.
 - This process is repeated until the best 10-indicator scorecard is constructed.

⁴³ The developers are the IRIS Center in the case of PAT and Mark Schreiner in the case of PPI.

An example of a finalized PPI scorecard and corresponding probability lookup table is included in Annex 4.

Tool construction: PAT

- **Indicator Selection.** The following steps are used to select the final set of indicators in PAT.
 - The developers first identify the highest quality, national survey data that contains information on expenditures or income. If a sufficiently recent or high-quality data set does not exist upon review, the developers may implement their own national survey. The total number of potential indicators is typically between 75 and 150.
 - Next, those indicators that are deemed “impractical” are discarded. From 2003 to 2006, IRIS carried out extensive consultation with MFI practitioners to identify indicators that were not practical, either because they were too difficult for respondents to answer, were easily falsifiable, or were sensitive to ask. This has been supplemented since 2006 by additional tool development experience and aggregated user feedback.
 - The developers have decided that all PATs will include indicators (control variables) for household size, household size squared, age of household head, age of household head squared, geographic location, and rural location.
 - The developers then use a statistical procedure called MaxR (short for maximization of R-squared) to select the additional 15 indicators that yield the best prediction of household expenditures. The MaxR procedure works as follows. In each iteration, a regression is run in which the dependent variable—either a continuous measure of household per capita expenditure or a binary poverty indicator based on expenditures—and the right-hand side variables are potential indicators. The iterative procedure for MaxR starts with a one indicator model (not including control variables), adds one variable per iteration to the regression model, and then considers all combinations among pairs of regressors to move from one iteration to the next until 15 variables are added to the model. The intuition is that, since R-squared gives the percentage of variation in households expenditures “explained” by the indicators, the higher the R-squared, the better the model’s ability to predict poverty status will be.
 - The MaxR procedure is run twice: once running OLS using the continuous household expenditure variable as the dependent variable and once running a linear probability regression using the binary poverty status as a dependent variable.⁴⁴ This results in two final sets of indicators (which may or may not be different).⁴⁵

⁴⁴ The MaxR procedure is not available for the quantile regression or probit regression.

⁴⁵ The PAT team also runs a two-step version of each of these four models. This approach is based on Grootaert and Braithwaite (1998). In the first step, the models are run over the full analysis sample. Those households who are predicted to be above a certain level of household expenditures are then excluded from the sample (as they are predicted

These two sets of indicators will then be used in the calculation of the weights associated with each indicator under four separate statistical models.

- **Calculation of weights.** The final indicators from these two MaxR procedures are then used in four separate statistical models.⁴⁶ They are:
 - *OLS Model* (explained above)
 - *Quantile Regression Model:* This model uses the final indicators identified using the MaxR procedure in OLS, which uses the continuous expenditure variable as the dependent variable. The weights are then based on the parameter estimates from the quantile regression.
 - *Linear Probability Model* (explained above)
 - *Probit Model:* This model uses the final set of indicators from the MaxR procedure in the linear probability regression, which uses the binary poverty status as the dependent variable. The weights are then based on the parameter estimates from the probit regression.
- **Selection of final model.** Once the final indicators are selected and associated weights are computed, the tool developers comparatively evaluate each of the eight models (including the four two-step models) based on accuracy. The model with highest accuracy is selected. In the next section (specifically in the subsection “Quality of Statistical Model”) we describe how accuracy is calculated for each tool. An example of a final PAT survey is included in Annex 2.

to be nonpoor) and the models run again on the households who remain in the data set. In addition to the control variables, 15 variables are selected in each step.

⁴⁶ PAT also uses a two-step procedure in which in the first step, OLS is used to identify the set of indicators. In the second step, weights are calculated by running the regression only on those households predicted poor in the first step.

ANNEX 8: Glossary and Explanation of Scoring

Additional user comments. User commentary on tool attributes (p. 4) that are not reflected in the Snapshot Review (p. 3). Comments represent a general summary of common responses.

Cost (tool attribute). Measured by a tool's financial cost, other requirements, and number of external consultants required to implement the tool.

Duration (tool attribute). Measures the time required to implement the assessment, divided into three components: training of staff and set up, survey time, and time to clean data.

Ease of use (tool attribute). Measures how the user perceives the burden of using a tool. Grading scale: quite difficult, somewhat difficult, somewhat easy, quite easy.

External consultants. Consultants (not employees of the MFI) working onsite for the purpose of implementing the assessment. For the purpose of the *User Reviews*, local surveyors that conduct client surveys are included in this definition.

Financial cost (tool attribute). Total monetary cost to implement the assessment. Grading: range of reported answers converted into U.S. dollars.

Level of confidence (tool attribute). Measures the degree to which a user felt the final score of the tool reflects the actual level of poverty of its clients. Grading scale: low, moderate, high.

Loan portfolio (US\$). All outstanding principal on all outstanding client loans, including current, delinquent, and restructured loans, but not loans that have been written off, as reported by Mix Market, an online MFI information data exchange.

MFI user requirements. The resources that an MFI must have in place, or be prepared to offer, in order to implement a tool.

Microfinance institution (MFI). Any institution that offers microfinance services.

MIS. Management information system.

Mixed (tool attribute rating). A rating used to indicate that user opinions were sufficiently varied so that no conclusion could be drawn.

Number of borrowers. The number of individuals who currently have an outstanding loan balance with an MFI or who are responsible for repaying any portion of the gross loan portfolio, as reported by Mix Market.

Number of external consultants (cost attribute). Number of external consultants who contracted to implement the assessment.

Offsite interviews. Interviews with MFI users of social performance tools that were conducted remotely by phone.

Onsite interviews. Interviews with MFI users of social performance tools that were conducted at the MFI's place of business.

Outreach data available? Indicates whether or not an MFI has social outreach data from the Progress out of Poverty Index (PPI), the Poverty Assessment Tool (PAT), or a credible impact evaluation tool.

Overall satisfaction (tool attribute). Measures whether or not a tool met an MFI's expectations and whether the institution was generally content with the process and results of the assessment. Grading scale: completely dissatisfied, mostly dissatisfied, mostly satisfied, completely satisfied.

Regions (of the world). For the purpose of the first round of *User Reviews* (2009), the four geographic regions of the world were Africa, Asia, Eastern Europe and the Middle East, and Latin America.

Regulatory status. The legal form of an MFI (e.g., bank, cooperative, nongovernmental organization, etc.).

Reported strengths/weaknesses. An MFI's perception of the strengths and weaknesses of the assessment process, as reported by a user. To be included in a *User Review*, a strength or weakness had to be cited by more than one user.

Robust MIS (tool requirement). Indicates whether or not an MFI has a reliable, flexible information system capable of tracking and reporting social performance indicators in such a way that external assessors have all the data required to complete an assessment or rating.

Robustness (tool attribute). Measures a user's perception of the accuracy of the tool and its results. Robustness is divided into three components: appropriateness of indicators, appropriateness of process, and level of confidence in final results.

Snapshot Review of Tool Attributes. Summary table of user comments on tool attributes (p. 3). See "Grading Methodology for Tool Attributes" (p. 17) for an explanation of how grading was conducted. *Note: while the "Snapshot Review" (p. 3) presents a summary of user responses, readers should consult "Additional User Comments" (p. 4) to get a more complete view of user feedback.*

Staff "champion." A staff member designated to work as part of a team, in conjunction with the external consultants performing a social assessment, to conduct interviews with stakeholders, compile results, and generate and present preliminary findings to institutional management. Staff champions typically receive training in tool methodology and may help write the final report.

Tool developer. The organization that developed a social performance assessment tool.

Tool outputs. The final outputs resulting from a social performance assessment or rating, which are included in the final report.

Usefulness (external). A tool attribute assessed by *User Reviews*. Measures a user's perception of the usefulness of assessment outputs for communicating or dealing with external entities, such as donors, government agencies, investors, network organizations, etc. Grading: useless, not very useful, useful, very useful.

Usefulness (internal), a tool attribute. Measures a user's perception of the usefulness of an assessment for internal purposes, such as planning, decision making, product development, etc.. Grading scale: useless, not very useful, useful, very useful.

Usefulness to MFI (tool attribute). Measures an MFI's perception of the utility of a given tool. Usefulness is divided into two components: internal and external (see above).

User. An MFI that has implemented a social performance assessment tool.

Annex 9: Explanation of Scoring

Grading Methodology for Tool Attributes

Tool attributes may be divided into two types: **quantitative** and **qualitative**.

For **quantitative attributes**, the reported range for all respondents is presented.

Example 1:

Four institutions were asked to report the financial cost of implementing poverty assessment tool “X”. Recorded answers were:

User:	#1	#2	#3	#4
Response:	\$12,000	\$15,000	Not Sure	\$9,000

Attribute rating: \$9,000–\$15,000.

Each **qualitative attribute** has its own grading scale, explained in the attribute definition in the Glossary. User responses for each attribute were compared to try to discern a general consensus “grade.” The following examples illustrate the process.

In some cases, a clear majority answer existed, and the given grade reflects this majority.

Example 2:

Four institutions were asked their opinion of the “ease of use” for poverty assessment tool “X”. Recorded answers were:

User:	#1	#2	#3	#4
Response:	High	High	Moderate	High

Attribute rating: High

In some cases, results were evenly split between two adjacent grades. The rating then reflects the two responses.

Example 3:

Four institutions were asked their opinion of the “ease of use” for poverty assessment tool “Y”. Recorded answers were:

User:	#1	#2	#3	#4
Response:	High	High	Moderate	Moderate

Attribute rating: Moderate to High

In some cases, responses were too varied to select a representative grade. Results were then recorded as “mixed.”

Example 4:

Six institutions were asked their opinion of the “ability to customize” social audit tool “C”. Recorded answers were:

User:	#1	#2	#3	#4
Response:	High	Low	Not Sure	Moderate

Attribute rating: Mixed

It is important to note that while “Snapshot Review of Tool Attributes” (p. 3) presents a summary of user responses, readers should consult “Additional User Comments” (p. 4) to get a more complete view of user feedback.

10. References

- Alatas, V., A. Banerjee, R. Hanna, B. Olken, and J. Tobias, J. 2010. "Targeting the Poor: Evidence from a Field Experiment in Indonesia." MIT, May. <http://econ-www.mit.edu/files/5577>
- Morduch, J. 2010. Targeting the Ultra Poor. Financial Access Initiative framing note. http://financialaccess.org/sites/default/files/RFN13_Targeting_The_Ultra_Poor%282010-8-10%29_0.pdf
- Ghalib, A. 2010. Does microfinance reach the poorest of the poor? Empirical evidence of programme outreach from rural Pakistan, IDPM, University of Manchester. http://www.chronicpoverty.org/uploads/publication_files/ghalib_microfinance_pakistan.pdf
- Banerjee, A., E. Duflo, R. Chattopadhyay, and J. Shapiro. 2007. "Targeting Efficiency: How Well Can We Identify the Poor?" Center for Microfinance, IFMR, December. http://www.povertyactionlab.org/sites/default/files/publications/89_Duflo_Banerjee_Targeting_Efficiency.pdf
- Chen, S., and M. Schreiner. 2009. "A Simple Poverty Scorecard for Bangladesh." <http://www.microfinance.com/#Bangladesh>.
- FINCA. 1998. "Poverty Assessment by Microfinance Institutions: A Review of Best Practice."
- Grootaert, C., and J. Braithwaite. 1998. "Poverty Correlates and Indicator-Based Targeting in Eastern Europe and the Former Soviet Union." World Bank Policy Research Working Paper No. 1942. <http://go.worldbank.org/VPMWVLU8E0>
- Schreiner, M. 2006. "Is One Simple Poverty Scorecard Enough for India?" http://www.microfinance.com/English/Papers/Scoring_Poverty_India_Segments.pdf
- Schreiner, M. 2009. "A Simple Poverty Scorecard for Peru." <http://www.progressoutofpoverty.org/peru>
- IRIS. 2005. "Developing and Testing Poverty Assessment Tools: Results from Accuracy Tests in Peru." Report prepared for USAID.
2010. "Pakistan's SSN Initiatives, Targeting Challenges & Suggestions." PowerPoint presentation. Washington, D.C.: World Bank. <http://siteresources.worldbank.org/SPLP/Resources/461653-1253133947335/6440424-1268060886824/Pakistan.pdf>
- Microfinance Poverty Assessment Tool. Henry, Sharmy, Lapenu and Zeller. Technical Tools Series No. 5. 2005
- Zeller, M. 2005. Review of Poverty Assessment Tools.