Certification Criteria: Accuracy
USAID/IRIS Tool for Guatemala
Submitted: October 31, 2006
Revised: November 9, 2006
May 10, 2013: Typos corrected

1. Please describe the overall approach to the tool development

The USAID/IRIS project on Developing Poverty Assessment Tools collected new data in four countries—Bangladesh, Peru, Kazakhstan, and Peru—to assess a selected set of indicators against the task of identifying "very poor" households based on household per capita expenditures. A composite survey questionnaire, compiled from several practitioner tools, was administered to a sample of 800 households. A benchmark for assessing measurement accuracy was developed using the expenditure module of the World Bank's *Living Standards Measurement Survey* (LSMS). Administered to the same set of households exactly fourteen days later, this benchmark provided the best available quantitative information on the "true" poverty status of each sample household. Multiple statistical methods were then used to identify the 15 indicators (for each "step" if a 2-step model) within this composite survey that most accurately reflect the "true" poverty status of each household – that is, that most closely track the benchmark expenditure results. Identification of the most accurate set of indicator and the weights attached to them were done of the basis of criteria developed especially for this project. In addition to the four countries already listed, a comparative analysis drew on existing LSMS data sets from an additional eight countries to identify the 15 best poverty predictors (using a similar methodology and set of variables), to facilitate generalization of findings over a larger number of countries. The eight LSMS countries are: Albania, Ghana, Guatemala, India (Bihar and Uttar Pradesh), Jamaica, Madagascar, Tajikistan, and Vietnam. Thus, statistical testing for accuracy was carried out for twelve countries in total. The 110 indicators that appeared in the 'best 15' from at least one of the twelve countries were included in the next part of the project: testing for practicality.

The 110 indicators were divided into six surveys to be tested for practicality. Seventeen microenterprise organizations were selected by USAID to conduct the field tests of practicality. Each question was rated as to whether the respondent found it to be sensitive, difficult, or that it was perceived that she falsified her answer. The lessons learned from the practicality testing were brought in after the best 15 poverty indicators were determined for each country. If a best 15 indicator caused difficulties in testing, the indicator was dropped for the list and the next best indicator replaced it.

The end result of this development process was a country-specific poverty assessment tool for each of the twelve countries that predicts—rather than directly measures—household per capita expenditure based on a short set of indicators. Each country tool is incorporated into a data entry template that allows microenterprise practitioner to easily enter and store the responses of its sampled clients to indicator questions and will also calculate the percentage of that practitioner's clients that are predicted to be very poor.

2. Please describe the data source used to develop or calibrate the tool.

Eight of the twelve country tools for this project were developed from existing LSMS data. For the other four countries, original survey data was collected, using both a composite survey consisting of poverty indicators from multiple sources and a benchmark expenditure survey based on the LSMS expenditure module. The sample was selected to be nationally representative.

For Guatemala, the poverty assessment tool was developed from a 2000 LSMS survey on a nationally-representative sample.[1] Economic growth (GDP per capita) in Guatemala from 2000 to present has averaged 0.2 percent per year, while inflation has averaged 7.4 percent per year.[2] While the data set for Guatemala is slightly more than five years old, we would argue that the alternative tool, based on a loan size, is likely to be far less accurate and thus a poor substitute.[3] The sample size used for the statistical testing of the tools was 7246 households.

3. Please describe the process used to select the indicators included in the tool.

LSMS data sets typically include variables related to education, housing characteristics, consumer durables, agricultural assets, financial assets, illness and disability, and employment. For Guatemala, roughly eighty indicators from all categories were considered.

The MAXR procedure in SAS was used to select the best poverty indicators (for variables found to be practical) from the pool of potential indicators in an automated manner. MAXR is commonly used to narrow a large pool of possible indicators into a more limited, yet statistically powerful set of indicators. The MAXR technique seeks to maximize explained variance (i.e., R^2) by adding one variable at a time (per step) to the regression model, and then considering all combinations among pairs of regressors to move from one step to the next. Thus, the MAXR technique allows us to identify the best model containing 15 variables (not including control variables for household size, age of the household head, and location).

The MAXR procedure yielded the best 15 variables for the OLS model (also used for the Quantile model) and another set of best 15 variables for the Linear Probability model (also used for the Probit model). The final set of indicators and their weights, therefore depended on selecting one of these four statistical models—OLS, Quantile, Linear Probability, or Probit—as the best model.[4] This selection of the best model was based on the BPAC and PIE accuracy criteria.

[1] Additional information on the data set is available here: http://www.worldbank.org/lsms/country/guat/gt00docs.html

[2] Data available from: http://www.imf.org/external/pubs/ft/weo/2006/01/data/dbginim.cfm

[3] When a loan size tool was tested in Bangladesh, its BPAC was strongly negative (-49.5). Please see: http://www.povertytools.org/Project_Documents/Bangladesh%20Accuracy%20Report%20Final.pdf

[4] The set of indicators and their weights also depended on the selection of a 1-step or 2-step statistical model.

4. Please describe the estimation methods used to identify final indicators and their weights/coefficients.

Guatemala (median) **Poverty Line:** 2680 quetzals per person per month[5] **Poverty Rate:** 22.96%	**Total Accuracy**	**Poverty Accuracy**	**Under-coverage**	**Leakage**	**PIE**	**BPAC**
Single-step methods						
OLS	87.02	68.47	31.53	25.03	-1.49	61.98
Quantile regression (estimation point: 44)	86.44	70.17	29.83	29.23	-0.14	69.56
Linear Probability	87.10	64.04	35.96	20.25	-3.60	48.34
Probit	87.99	70.33	29.67	22.62	-1.62	63.28
Two-step methods						
OLS – 40 percentile cutoff	88.03	72.66	27.35	24.79	-0.59	70.09
Quantile (estimation points: 44, 19) 40 percentile cutoff	87.96	74.25	25.75	26.72	0.22	73.28
LP – 35 percentile cutoff	88.39	74.70	25.29	25.27	-0.01	74.67
Probit – 35 percentile cutoff	88.21	73.81	26.19	25.19	-0.23	72.79

	Number of households[6] predicted as very poor by the tool	**Number of households predicted as not very-poor by the tool**
Number of "true" very poor households (as determined by benchmark survey)	1242 (17.14%)	421 (5.81%)
Number of "true" not very-poor households (as determined by benchmark survey)	420 (5.80%)	5163 (71.25%)

[5] The poverty line reported above is expressed in the prevailing prices at the time of data collection (November 2000). The poverty line and other monetary variables used in the accuracy results, however, are inflation-adjusted through the time of tool construction (June 2006), to ensure that the resulting poverty predictions are as current as possible. The poverty line for Guatemala, expressed in the most current available prices, is 48,008.39 quetzals per person per year.

[6] Values in the matrix are weighted for national representation.

5. Please describe how coefficients and weights are used to compute prediction of poverty status or estimate of household expenditures.

The weights attached to the indicators in the tool in each country are simply the regression coefficients for the statistical model exhibiting the highest statistical accuracy (according to the BPAC criterion). For Guatemala, the weights are from a 2-step Linear Probability model.

The weights are located in the "backend" analysis program of the EPI template as part of the extreme poverty rate calculation. While a skilled EPI user would be able to locate these values and potentially manipulate them, they would not be seen by the client or the interviewer during the normal course of interviewing, entering the data, or in calculating the extreme poverty rate.

6. Please describe the decision rule used to classify households as very poor or not very-poor.

The extreme poverty line for each of the twelve countries in the project was the higher of the two potential poverty lines specified in the legislation: $1.08 a day (in PPP terms) OR the bottom half of households living below the poverty line (termed the 'median' poverty line). Two of statistical models (OLS and Quantile) used by the IRIS team predict the per capita consumption expenditures for each household, which is then compared to the binding poverty line to decide whether the household is very poor.[7] The other two statistical models (Linear Probability and Probit) predict the probability that a household is very poor (according to the binding, absolute poverty line). If this probability exceeds 0.5, the household is predicted to be very poor.[8]

For Guatemala, the binding poverty line was the 'median' line of 48,008.39 quetzals per person per year. Because the tools is based on a Linear Probability model, those households whose probability of having expenditures below this line is greater than 0.5 will be considered very poor.

[7] For a 2-step OLS or Quantile model, the decision rule in the 1st-step compares the expenditures predicted for each household to a certain expenditure cutoff.

[8] For a 2-step Linear Probability or Probit model, the decision rule in the 1st-step compares the predicted probability that the households' expenditures exceed a certain cutoff to the 0.5 value.

GUATEMALA 2 STEP MAXR/LP: 100 percentile
The REG Procedure
Number of Observations Used: 7246

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	27	830.22972	30.74925	271.00	<.0001
Error	7218	818.99840	0.11347		
Corrected Total	7245	1649.22812			

Root MSE	0.33685	R-Square	0.5034
Dependent Mean	0.64965	Adj R-Sq	0.5015

Variable	Coeff.	Std. Err.	t	P > |t|	Standardized Estimate
Intercept	0.7849	0.0410	19.13	<.0001	0.0000
Household size	-0.1201	0.0058	-20.87	<.0001	-0.6356
Household size squared	0.0045	0.0004	10.40	<.0001	0.3047
Household head age	0.0041	0.0016	2.56	0.0105	0.1287
Household head age squared	-0.0001	0.0000	-2.48	0.0131	-0.1244
Household lives in the Capital	0.0483	0.0150	3.21	0.0013	0.0438
Household lives in the North	-0.1132	0.0199	-5.69	<.0001	-0.0614
Household lives in the Northeast	0.0378	0.0182	2.08	0.0379	0.0225
Household lives in the Southeast	-0.0102	0.0187	-0.55	0.5848	-0.0060
Household lives in the Southwest	-0.0238	0.0145	-1.64	0.1012	-0.0216
Household lives in the Northwest	-0.0362	0.0180	-2.01	0.0443	-0.0239
Household lives in Peten	0.0410	0.0264	1.55	0.1213	0.0146
Household lives in rural area	-0.0353	0.0110	-3.2	0.0014	-0.0367
Household head is literate in Spanish	0.0484	0.0099	4.87	<.0001	0.0472
Share of household members who have incomplete primary as their highest level of education *	0.1470	0.0222	6.62	<.0001	0.0612
Share of household members who completed primary as their highest level of education*	0.2726	0.0344	7.93	<.0001	0.0703
Share of household members who have incomplete secondary as their highest level of education*	0.2989	0.0334	8.95	<.0001	0.084
Number of days of work lost by household members to illness or accident	0.0061	0.0015	4.17	<.0001	0.0360
Wall is made of adobe or wattle and daub	-0.0667	0.0100	-6.65	<.0001	-0.0627
Household did not use electricity in the past month	-0.0718	0.0119	-6.06	<.0001	-0.0666
Household does not have sanitary system	-0.0639	0.0129	-4.94	<.0001	-0.0453
Household owns one or more mixers	0.0722	0.0116	6.24	<.0001	0.0736
Household owns one or more pick ups	0.0860	0.0172	4.99	<.0001	0.0431
Household owns one or more televisions	0.1088	0.0115	9.44	<.0001	0.1136
Household owns one or more stoves	0.1754	0.0126	13.94	<.0001	0.18373
Household owns one or more bicycles	0.0658	0.0092	7.15	<.0001	0.06487
Household owns one or more agricultural mills	0.2134	0.0395	5.4	<.0001	0.0454
Number of pigs owned by household	0.0144	0.0032	4.53	<.0001	0.03922

*Does not include household head

MAXR/LP on sample below cutoff of 35 percentiles
The REG Procedure
Number of Observations Used: 2635

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	27	158.22460	5.86017	31.75	<.0001
Error	2607	481.12449	0.18455		
Corrected Total	2634	639.34909			

Root MSE	0.42959	R-Square	0.2475
Dependent Mean	0.58194	Adj R-Sq	0.2397
Coeff Var	73.82069		

Variable	Coeff.	Std. Err.	t	P > |t|	Standardized Estimate
Intercept	-0.2373	0.0985	-2.41	0.0161	0.0000
Household size	0.2058	0.0148	13.94	<.0001	1.0309
Household size squared	-0.0084	0.0009	-8.90	<.0001	-0.6391
Household head age	-0.0060	0.0038	-1.58	0.1134	-0.1696
Household head age squared	0.0001	.0000	1.48	0.1386	0.1570
Household lives in the Capital	-0.3986	0.0560	-7.11	<.0001	-0.1475
Household lives in the North	0.0630	0.0393	1.60	0.1089	0.0465
Household lives in the Northeast	-0.1460	0.0462	-3.16	0.0016	-0.0754
Household lives in the Southeast	0.0174	0.0420	0.41	0.6786	0.0111
Household lives in the Southwest	0.0358	0.0358	1.00	0.3165	0.0330
Household lives in the Northwest	0.0126	0.0378	0.33	0.7397	0.0107
Household lives in Peten	-0.1070	0.0536	-2.00	0.0460	-0.0426
Household lives in rural area	0.0694	0.0287	2.42	0.0156	0.0432
Household head has complete primary as highest level of education	-0.1368	0.0354	-3.87	0.0001	-0.0678
Share of household members who have incomplete primary as their highest level of education *	-0.3038	0.0489	-6.21	<.0001	-0.1151
Share of household members who completed primary as their highest level of education*	-0.7799	0.1125	-6.94	<.0001	-0.1232
Share of household members who have incomplete secondary as their highest level of education*	-0.7495	0.1933	-3.88	0.0001	-0.0686
Roof is made of tiles, shingles or other materials	-0.0850	0.0239	-3.55	0.0004	-0.0689
Wall is made of adobe or wattle and daub	0.1354	0.0211	6.43	<.0001	0.1370
Wall is made of cane or sticks or other materials	0.1049	0.0283	3.71	0.0002	0.0719
Household did not use electricity in the past month	0.0962	0.0192	5.02	<.0001	0.0966
Household owns one or more landline telephones	-0.3120	0.0891	-3.50	0.0005	-0.0612
Household owns one or more cameras	-0.3031	0.0721	-4.20	<.0001	-0.0728
Household owns one or more cars	-0.3728	0.0855	-4.36	<.0001	-0.0758
Household owns one or more bicycles	-0.0902	0.0230	-3.92	<.0001	-0.0690
Household owns one or more sewing machines	-0.1662	0.0388	-4.29	<.0001	-0.0753
Number of pigs owned by household	-0.0296	0.0066	-4.47	<.0001	-0.0783

Household had a savings account (term deposit, ordinary savings, or monetary deposits)	-0.2203	0.0582	-3.78	0.0002	-0.0652

*Does not include household head