WPS7182

Policy Research Working Paper 7182

Same Question but Different Answer

Experimental Evidence on Questionnaire Design's Impact on Poverty Measured by Proxies

Talip Kilic
Thomas Sohnesen

WORLD BANK GROUP
Development Research Group
Surveys and Methods Team
January 2015

Abstract

Does the same question asked of the same population yield the same answer in face-to-face interviews when other parts of the questionnaire are altered? If not, what would be the implications for proxy-based poverty measurement? Relying on a randomized household survey experiment implemented in Malawi, this study finds that observationally equivalent as well as same households answer the same questions differently when interviewed with a short questionnaire versus the longer counterpart that, in a prior survey round, would have informed the prediction model for a proxy-based poverty measurement exercise. The analysis yields statistically significant differences in reporting between the short and long questionnaires across all topics and types of questions. The reporting differences result in significantly different predicted poverty rates and Gini coefficients. While the difference in predictions ranges from approximately 3 to 7 percentage points depending on the model specification, restricting the proxies to those collected prior the variation in questionnaire design, namely demographic variables from the household roster and location fixed effects, leads to same predictions in both samples. The findings emphasize the need for further methodological research, and suggest that short questionnaires designed for proxy-based poverty measurement should be piloted, prior to implementation, in parallel with the longer questionnaire from which they have evolved. The fact that at the median it took 25 minutes to complete the food and non-food consumption sections in the long questionnaire also implies that the implementation of these sections might not be as overly costly as usually assumed.

This paper is a product of the Surveys and Methods Team, Development Research Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at tkilic@worldbank.org.

Same Question but Different Answer: Experimental Evidence on Questionnaire Design's Impact on Poverty Measured by Proxies

Talip Kilic and Thomas Sohnesen[1]

Keywords: Poverty Measurement, Proxy-Based Poverty Measurement, Survey-to-Survey Imputation, Questionnaire Design, Household Survey Experiment, Malawi, Sub-Saharan Africa.
JEL Codes: C83, D12, I32.

---

[1] The affiliation of the authors: Talip Kilic: Senior Economist, Living Standards Measurement Study, Surveys and Methods Group, Development Research Group, The World Bank, 1818 H St. NW, Washington, DC, 20433 USA. Thomas Pave Sohnesen: Living Standards Measurement Study, Surveys and Methods Group, Development Research Group, The World Bank, 1818 H St. NW, Washington, DC, 20433 USA, and Visiting Professor with the Development Economics Research Group, University of Copenhagen, Denmark. The corresponding author: Talip Kilic; tkilic@worldbank.org; +1-202-458-5892. The authors would like to thank Heather G. Moylan for excellent research assistance; Phyllis Ronek for editorial assistance; Alberto Zezza, Astrid Mathiassen and the World Bank Poverty and Inequality Measurement and Analysis Practice Group members for their comments; and Astrid Mathiassen and Bjorn Wold for their inputs into the experiment design.

1. INTRODUCTION

Does the same question that is asked of the same population yield the same answer in face-to-face interviews when other parts of the questionnaire are altered? If not, what might be correlated with the discrepancies and what would be the resulting implications for proxy-based poverty measurement? These are the central questions of our study. While the empirical investigation is conducted in the context of predicting household consumption expenditures, our findings are equally relevant for the estimation of trends based on questionnaires that exhibit variations in design over time and for impact evaluations that might rely on questionnaires of different length and complexity for treatment and control samples.

Estimating consumption and poverty via proxies is compelling as consumption measurement is often argued to be complex and costly. The literature on proxy-based poverty measurement highlights the promise of the method in improving the frequency and comparability of inter-annual poverty estimates at a lower cost, and sometimes by using already existing data (Christiaensen et al., 2012). Related applications on providing intra-year poverty predictions (Douidich et al., 2013), and developing proxy means tests for enhancing the targeting performance of development programs (Houssou & Zeller, 2011) have captured attention. With increasing pressure placed on national statistical systems by governments and the international community for increasing the frequency, quality, and comparability of poverty statistics, the interest in the method's application for filling the gaps in a cost effective fashion is generating continued interest.

Both parametric and non-parametric approaches to estimation have been featured in the literature.[2] Regardless of the approach, all practical applications would rely on data originating from two non-identical questionnaires: one set of data to establish the underlying model and another set of data with proxies to pair with the model parameters and to obtain predictions. In the case of consumption and poverty, the model is typically established based on data from a multi-purpose household questionnaire that yields a comprehensive welfare aggregate (hereafter referred to as a standard household questionnaire), while data on proxies would be solicited through a shorter household questionnaire, often with a shorter field implementation period.[3] Even if questions underlying proxy definitions are worded *identically* across short versus standard household questionnaires, identical questions could yield different answers in questionnaires that exhibit substantial variation in inter- and intra-module scope of data collection.

[2] Vu and Baulch (2011) evaluate a range of these methods in the context of Vietnam.

[3] Although the shorter fieldwork for a short household survey would result in cost savings, the differences in the period of implementation between a short household survey and its standard comparator could affect the values obtained for the seasonality-prone poverty proxies. Our set-up ensures that the observed differences between the data obtained from a short vs. standard questionnaire are not due to differences in the period of survey implementation.

The interactions between questionnaire design and cognitive processes underlying reporting are complex. Tourangeau et al. (2000) posit that question answering process involves the stages of comprehension, retrieval, judgment, and response production. Theoretically, questionnaire design decisions place different demands on respondents at different stages (Hess et al., 2001). The literature on substantial questionnaire variation with identical questions given to comparable samples is thin. Beegle et al. (2012), in their comparative assessment of methods of household consumption in Tanzania, find that the recall-based reporting on frequent non-food consumption expenditures is negatively affected by increasing the scope of the food consumption module (whether recall- or diary-based) which is administered prior to the non-food consumption module. Given the questionnaire wording and structure for the non-food consumption module was identical across the food consumption module variants, the authors suggest respondent burden to be the potential culprit behind their finding.[4] The evidence on the presence of respondent burden and its effects on data quality are quite heterogeneous. The documented effects are ultimately context- and subject-specific, but there are several (some experimental) studies that document (i) question/module placement effects, whether earlier or later in a questionnaire (Johnson et al., 1974; Kraut et al., 1975; Herzog & Bachman, 1981; Andrews, 1984), and (ii) motivational under-reporting during personal interviews in responses to gateway questions as to avoid follow-up questions (Kreuter et al., 2011; Eckman et al., forthcoming).[5]

If independent samples that are drawn from the same population and that are interviewed at the same time, indeed provide different values for the same poverty proxies depending on whether they were subject to a short questionnaire versus its standard counterpart that would be used for establishing the poverty prediction model in a prior period, it is reasonable to expect that the subsequent poverty predictions could be different.[6] Our study is the first that provides experimental evidence on this possibility, which is implicitly assumed away in proxy-based poverty measurement exercises as long as questions underlying proxy definitions are worded identically across short and standard survey instruments.[7]

[4] Though not reported in their paper, Beegle et al. (2012) experimented also with the placement of the labor module of the questionnaire, which was put randomly either before or after in 4 of their 8 consumption designs. They looked at whether food as well as total consumption were impacted in each of these four scenarios (i.e. 8 regressions) and found mixed results, with a modest suggestion that both food and total consumption were lower when the labor module came before (i.e. statistically significant effect at 10 percent level in 2 of the 8 regressions). These insights were obtained in private communication with the authors. The experimentation around survey design in their study is not as comprehensive as a shift from a standard to a light household survey would typically be, and the observed impacts could therefore be different.

[5] The data collection themes across the cited studies do not overlap with those featured in our analysis.

[6] Newhouse et al. (2014) document a Sri Lankan application in which proxy-based poverty predictions fail to track official poverty estimates. In the urban areas, they identify the cause of the problem as the incomparability of the employment *question* between their light and standard household questionnaires. Even though our study highlights the potential incomparability of the *data* by light vs. standard questionnaire treatment, Newhouse et al. (2014) highlight another type of sensitivity of proxy-based poverty estimation to changes in questionnaire design.

[7] Survey mode does not differ between the light and standard household questionnaires used in our experiment, and we rely on paper questionnaires administered by interviewers in face-to-face interviews. There is a rich literature on the comparative effects of survey mode (computer-assisted personal interviewing in face-to-face interviews,

More specifically, the work is based on a randomized household survey experiment that was implemented in Malawi in 2013. The inspiration for the experiment was the discrepancy in the poverty trends based on competing Malawi National Statistical Office (NSO) products during the period of 2004/05-2010/11. Although the direct measurement of household consumption expenditures from the Second Integrated Household Survey (IHS2) and the Third Integrated Household Survey (IHS3) had produced a stagnant headcount poverty trend of 52 percent in 2004/05 and 51 percent in 2010/11, the Welfare Monitoring Survey (WMS)-based poverty predictions that were disseminated between the IHS2 and the IHS3 had implied a steep decline from 50 percent in 2005 to 39 percent in 2009. At conceptualization, the WMS had been designed to provide, among other indicators, poverty predictions on an annual basis in the interim years of the IHS, which is conducted approximately every 5 years. This objective was fulfilled between the IHS2 and the IHS3 by combining the parameters from a model of household consumption expenditures estimated using the IHS2 with the associated proxies obtained from the 20-page WMS questionnaire that was markedly lighter in inter- and intra-module scope of data collection than the IHS counterpart.[8]

There are three key findings that emerge from the analysis. First, we find that observationally equivalent households as well as same households answer the same questions differently when interviewed with a short questionnaire versus its standard counterpart. Second, the analysis yields statistically significant differences in reporting across all topics and types of questions. The effect is quite pronounced for binary poverty proxies related to consumption of non-food and food consumption items, and experience of household shocks. The categorical variables, particularly those related to subjective welfare and housing, are also impacted by changes in questionnaire design. Third, relying on prediction models based on the national household survey data collected with the standard questionnaire in 2010, we find that the differences in reporting are sufficient to give poverty predictions that are significantly different from each other. While the resulting difference in predicted poverty estimates ranges from approximately 3 to 7 percentage points depending on the model specification, restricting the poverty proxies to the ones that do not differ by survey treatment, namely demographic variables from the household roster and location fixed effects, predicts same poverty rates in both samples. The findings emphasize the need for further methodological research on module/question placement effects and associated cognitive and behavioral processes, and support the view that light household survey operations designed for proxy-based poverty measurement should judiciously pilot their instruments prior to roll-out, in parallel with the questionnaire instruments from which they have evolved.

---

telephone interviews, self-administered questionnaires mailed-in or completed online, etc) that we do not delve into. If survey mode differs between the light and standard household questionnaires used in a proxy-based poverty measurement exercise, the variation may affect the proxy measurement and the subsequent poverty predictions.

[8] The information on the WMS is available on http://www.nsomalawi.mw/publications/welfare-monitoring-surveys-wms.html.

The paper is organized as follows. Section 2 presents the randomized household survey experiment set up and describes the data. Section 3 assesses the differences in reporting by survey treatment status and reports the findings. Section 4 evaluates the impact on proxy-based poverty measurement. Section 5 concludes.

2. DATA

The methodological experiment on proxy-based poverty measurement (hereafter referred to as "the experiment") was integrated into the Malawi Integrated Household Panel Survey (IHPS) 2013, which was implemented using paper questionnaires and face-to-face interviews. The IHPS attempted to track and resurvey 3,246 households across 204 enumeration areas (EAs) that had been surveyed for the Third Integrated Household Survey (IHS3) 2010/11.[9] The survey was implemented by the National Statistical Office (NSO), and had been designed at baseline to be representative at the national-, urban/rural, regional levels, and for the six strata defined by the combinations of region and urban/rural domains. The IHPS targeted all individuals that were part of the IHS3, including those that moved away from the IHS3 dwelling locations between 2010 and 2013. Once a split-off individual was located, the new household that he/she formed or joined since the IHS3 interview was brought into the IHPS sample. As a result, the overall IHPS database includes 4,000 households, which could be traced back to 3,104 IHS3 households.[10,11]

The main IHPS fieldwork was carried out during the period of April-October 2013, with residual tracking operations conducted during the period of November-December 2013. The survey attempted to visit each household twice, identical to the IHS3 practice, within two weeks of the baseline interview timeframe, and with approximately three months in between visits. At baseline, the IHPS EAs had been randomly divided into two halves, known as Sample A and Sample B EAs, and the questionnaire load for households in these EAs had been split differently across visits. During the IHPS, Sample A households were administered the standard household questionnaire during Visit 1, and had simply received an update to the household roster module in Visit 2. In contrast, Sample B households had received only the household roster module of

[9] The IHPS 2013 and the IHS3 were supported by the Living Standards Measurement Study – Integrated Surveys on Agriculture (LSMS-ISA) initiative, whose primary objective is to provide financial and technical support to governments in sub-Saharan Africa in the design and implementation of nationally-representative multi-topic panel household surveys with a strong focus on agriculture. The IHPS 2013 and IHS3 data and documentation are publically available on www.worldbank.org/lsms.
[10] Attrition was limited to only 3.78 and 7.42 percent, at household and individual levels, respectively.
[11] The interviewer training for the experiment was part of the IHPS field staff training; as such it was hands-on and extensive. During the fieldwork, the field staff was under continuous quality control. The threat of (and actual) supervisor re-interviews of households and systematic interviewer evaluations by the IHPS management attempted to prevent interviewer-specific tendencies from culminating. This was reinforced by ensuring in each half of the fieldwork that the experiment workload was spread evenly across the interviewers, alongside their main IHPS assignments. The experiment data processing and quality control protocols were identical to the IHPS protocols.

the standard household questionnaire in Visit 1 and had been administered the rest of the questionnaire in Visit 2.

The standard household questionnaire spanned 66 pages and 23 modules. Our experiment was administering an additional 2-page instrument (included in the Appendix), immediately after the household roster module (i.e. the first module following the cover page), to a subsample of IHPS households during the visit in which the interview would have only necessitated the administration of the household roster module. Toward this end, 4 households in each IHPS EA, out of the households that remained in the original EA between 2010 and 2013, were randomly selected for the experiment, and received the additional 2-page instrument. Since only households that had remained in the original EA were considered for the experiment, we limit the analysis sample to all households that remained in the original EA between 2010 and 2013, and that were subject to the two-visit approach in 2013.[12] This yields an analysis sample of 2,822 households, out of which 765 households were part of the experiment.[13]

Table 1 provides an overview of the sample. Out of 1,428 Sample A households, who received the full standard questionnaire in Visit 1 and were revisited in Visit 2 for a household roster update, 393 households also received the additional 2-page instrument following the household roster module in Visit 2. Similarly, out of 1,394 Sample B households, who received only the household roster module in Visit 1 and the full standard questionnaire in Visit 2, 372 households were administered the additional 2-page instrument in Visit 1. Hence, 765 experiment records form a sub-sample of whom the same questions were asked in different questionnaires and at different points in time.

In selecting the questions to be included in the 2-page instrument for the experiment, we solicited inputs from the Statistics Norway staff that had supported the NSO in producing WMS-based poverty predictions, and aimed to (i) be able to compute the indicators from the Poverty Predictors module of the WMS questionnaire[14], (ii) capture the poverty proxies used by past survey-to-survey imputation applications to the Malawi Second Integrated Household Survey (IHS2) 2004/05 data (Houssou & Zeller, 2011), and (iii) include other poverty proxies on food consumption, non-food consumption and subjective welfare that have been suggested in the literature but that are not currently used extensively (Christiaensen et al., 2012).

[12] Given the demanding tracking objectives of the survey, the teams managed to implement the two-visit approach for 91.7 percent of the IHPS sample (i.e. 3,667 households). On average, there were 96 days between the two visits.

[13] Table A1 in the Appendix presents the sample means for 36 household level attributes computed from the non-experiment modules and the results from the tests of mean differences by whether a household was part of the experiment. No mean difference is statistically significant at least at the 10 percent level, supporting the view that on average, there are no systematic differences between the non-experiment and the experiment samples beyond the difference in the survey treatment that they were subject to randomly.

[14] The Poverty Predictors module of the WMS questionnaire was unchanged during the period of 2005-2009, and also consisted of 2 pages.

The modules that were part of the 2-page instrument were abbreviated versions of the following modules in the standard household questionnaire: (i) housing, (ii) food consumption over past one week, (iii) non-food expenditures over past one week and one month, (iv) non-food expenditures over past three months, (v) durable goods, (vi) shocks and coping strategies, and (vii) subjective assessment of well-being. Since the durable goods module was inadvertently different across survey treatments, we elect not to use the data on the ownership of durable goods in our analysis.[15] The modules were administered in the same order in which they appeared in the standard household questionnaire[16] and yield a mix of binary (70), ordered categorical (12), and continuous (1) poverty proxies.

Table 2 presents the median time allocated to the administration of a given module, and the median time the interview had been on-going prior to the administration of the module in question. The statistics are presented separately for the experiment and standard interviews. The complexity and scope of the standard household questionnaire lead to substantially longer interviews. By the time the first poverty proxy question is asked in the standard interview (at the 34th minute mark at the median), the experiment interview is already conducted in full (at the 23rd minute park at the median). Another striking finding is that the standard questionnaire modules on food and non-food consumption that we seek to proxy in fact take less than 25 minutes to administer as a package at the median. This brings into question, at least in the case of Malawi, why consumption data, in and of itself, is considered complex and too costly to collect at a higher frequency.

3. REPORTING DIFFERENCES IN EXPERIMENT VS. STANDARD INTERVIEWS

To explore reporting differences by household survey treatment status, we estimate multivariate regressions of the following form:

$$y_i = \alpha + \beta e_i + \gamma Z_i + \mu_i \qquad (1)$$

where i stands for household; y is a binary, categorical or continuous poverty proxy of interest; e is the binary variable identifying whether or not a household was part of the experiment; Z *is* a

[15] In the IHPS household questionnaire, the ownership of each asset is first established by a yes/no question, with the values of 1 and 2 recorded for yes and no answers, respectively. The question on the number of items owned is then asked for assets that are owned. Due to a mistake in the design of the experiment instrument, the yes/no question was dropped, and the question on the number of items owned was included with an instruction for the interviewer to record a value of zero for assets that are not owned. This resulted in an unusual number of experiment households owning two assets in the Visit 1 data, which led to the discovery of the fact that interviewers were recording a value of 2 in the experiment module for assets that are not owned, similar to the practice followed for the yes/no question in the complex household questionnaire. Although the interviewers were retrained on the correct administration of the experiment module prior to the Visit 2 period, we still do not have 100 percent confidence in these data.

[16] The only exceptions were (vi) and (vii) whose order was reversed in the 2-page instrument for presentation reasons.

vector of observable household attributes computed from non-experiment modules; α and μ identify constant and error terms, respectively.[17] For binary, categorical and continuous poverty proxies, we use Logit, Ordered Logit, and OLS regressions, respectively.

In what follows, we follow two approaches in estimating the survey treatment effect, β. The first relies on the comparison of poverty proxy values reported by 2,057 non-experiment households during their standard interviews vis-à-vis the values reported for the same outcomes by 765 experiment households during their short interviews. The results from this line of analysis are reported under the "*Experiment versus Standard*" heading in the tables that follow. The second line of analysis attempts to gauge the sensitivity of these findings by comparing the answers provided by the 765 experiment households during their standard vs. experiment interviews. The tables report the results from this line of analysis are reported under the "*Same Households: Experiment versus Standard*" heading.

There are significant discrepancies in how households answer the same questions in different questionnaires.[18] Table 3 reports module-specific counts of poverty proxies that are associated with statistically significant survey treatment effects at least at 10 percent level. Comparing the experiment versus standard samples, there are significant differences in 33 of the 83 variables (column 2), equivalent to significant differences in approximately 40 percent of the variables. Even when you ask the same households the same questions, the answers often differ. In 32 of the 83 questions, equivalent to 39 percent of the questions, we get significantly different answers for the same questions among households that were asked these questions in different questionnaires at two different times (column 3). The fact that same households also answer the same questions differently depending on the questionnaire instrument is strong evidence that the variation in the questionnaire design is driving these results.

The only continuous poverty proxy, namely household cell phone expenditures, has an average that differs between the experiment and standard samples but not when the experiment versus

[17] Given the evidence for successful randomization, bivariate statistical tests should theoretically provide sufficient evidence for whether reporting differences exist by household survey treatment status. The vector of controls included in Equation (1), however, allows us to account for any remaining unobservable heterogeneity correlated with the observed attributes and to explore heterogeneity of impact later in the analysis. The results are indeed not sensitive to whether or not the vector of controls is included in Equation (1). The vector of control variables include (i) household size and sum of household members aged 0-14 and over the age of 65; (ii) age (in years) of head of household, (iii) binary variable identifying female head of households; (iv) binary variables identifying the highest educational attainment among household members, capturing primary, junior secondary, and secondary (and above) educational attainment; (v) binary variables identifying Christian and Muslim head of households, (vi) binary variables identifying Chewa and Tumbuka head of households; (vii) binary variables capturing polygamous, separated, divorced, widowed/widower, never married head of households, (viii) number of months in the last 12 months that head of household has been away; (ix) number of days in the last 7 days that head of household has been away; (x) binary variable identifying rural/urban residence, (xi) binary variables capturing north and south regional location, and (xii) month of interview fixed effects.

[18] There are no differences in item non-response among different samples of interest. On the whole, the item non-response is present only in 0.02 percent across all comparable questions in each sample.

standard interview values are compared for the same experiment households. The shares of categorical versus binary poverty proxies exhibiting significant survey treatment effects are comparable regardless of the sample specification. Among housing variables, we observe a consistent significant difference only in reporting for the toilet type across the sample comparisons.[19]

The four questions on subjective well-being include three questions asking households to place themselves, their friends, and their neighbors on a six-point scale going from poor to rich, and a question asking households if they find their consumption less than, equal to, or more than adequate. We find a significant difference between the experiment and standard samples in how they rate the welfare level of their friends (column 2) and a significant difference in how the experiment households rate their own welfare, depending on whether they are subject to an experiment versus standard interview (column 3). Regarding the four ordinal categorical questions on durable assets (i.e., number of bednets in the household, number of phones in the household, sets of clothing for the head of household and the quality of bed sheets for the head of household), we observe a significant difference for the quality of bed sheets for the head of household (columns 2 and 3).

For proxy-based consumption and poverty measurement, it matters greatly, if the differences in reporting are systematic. Although not shown in Table 3 explicitly, out of the 27 binary outcomes that exhibit statistically significant differences in column 2, 24 of them have a higher mean in the experiment sample. To investigate this pattern further, we pool all binary poverty proxies, and, separately, all ordered categorical poverty proxies (but in batches, in accordance with the number of categories), and use the resulting pooled data in estimating Equation 1. Table 4 presents from these estimations the marginal effect and standard error associated with the binary variable that identifies whether a household was subject to the experiment. We rely on Logit and Ordered Logit regressions for the analysis of binary and ordered categorical outcomes, respectively. For Ordered Logit estimations, we report the marginal effect on the probability of being in the lowest category.[20] We present results without controls in column 1 and 4 and with controls, as specified in Equation 1, in columns 2, 3, 5 and 6. The results are robust varying the scope of the control variables (both those included and other alternatives) and the sample comparisons.

The core results reported in columns 2 and 5 indicate that the experiment questionnaire treatment, on average, translates into 2.3 percentage point increase in the probability of a positive answer for binary poverty proxies. At the mean of 25.7 percentage points for the standard sample, this effect is equivalent to 8.9 percent higher reporting. Regarding the ordered

[19] Other housing attributes include the roof and floor type and the number of rooms in dwelling. The roof and floor type are typically assessed by enumerators, without asking the household.

[20] The full spectrum of marginal effects estimated in each category of each pooled ordered categorical variable set are reported in the Appendix Tables A2 through A5.

categorical variables, much of the traction is in the analysis of outcomes with 3 and 6 categories. The marginal effect of experiment questionnaire treatment on the probability of being in the lowest category for the pooled categorical poverty proxies with 3 categories, on average, ranges from 1.8 to 2.0 percent. Similarly, the marginal effect of short questionnaire treatment on the probability of being in the lowest category for the pooled categorical outcomes with 6 categories (all originating from the subjective well-being module), on average, ranges from 2.3 to 3.2 percent.

3.1. HETEROGENEITY IN REPORTING DIFFERENCES

The systematic higher reporting associated with binary poverty proxies in the experiment sample is likely to result in systematic different estimation of consumption and poverty. However, is the impact equal for all modules and comparison groups? The results in Table 5 shed light on this aspect of heterogeneity of the experiment questionnaire treatment impact on binary poverty proxies. The results are not sensitive to (i) using the data from either the first or the second half of the fieldwork or (ii) focusing on the sample of households that received the same questions in different questionnaires at two different points in time (columns 2, 4, 6 and 8). We note a larger effect for the binary poverty proxies capturing the consumption of non-food items and a smaller effect for those capturing the experience of shocks in the last 12 months. Evaluating the coefficients reported in columns 3, 5 and 7 in the context of the mean from the corresponding module in the standard sample, we observe that experiment questionnaire treatment corresponds to higher reporting in the amount of 7.1 percent for food consumption, 12.4 percent for non-food consumption, and 7.9 percent for experience of shocks.

Of interest is also whether the experiment questionnaire treatment effect varies with household attributes. If it does, the predictions based on poverty proxies that are not immune to the experiment questionnaire treatment are likely to result in a different shape of the consumption distribution, as opposed to a mere level effect. To shed light on this possibility, we estimate Logit regressions using the pooled binary poverty proxies and interact the experiment questionnaire treatment identifier with selected household attributes, while controlling for the vector of control variables. Table 6 reports the findings from this analysis, which is based on the comparison of the experiment and standard samples.[21]

Households characterized by being larger and residing in urban areas, are, on average, more likely to answer yes to questions on both food and non-food consumption when interviewed with the experiment questionnaire (columns 2 and 3). As the number of dependents decline and the

[21] The results based on the comparisons of the experiment versus standard interview values of poverty proxies for the experiment households are near-identical to the patterns in table 6, and are not reported in the interest of brevity. They are, however, available upon request.

household is subject to the experiment questionnaire treatment, the likelihood of reporting positive non-food consumption also increases. The household attributes that are underlining the statistically significant interaction effects are commonly associated with richer households, whose higher likelihood of reporting different answers to same questions in different questionnaires is likely partly "mechanical". Since these households also consume more, they also have more scope for answering differently. On the other hand, the experiment questionnaire treatment effect on the reporting of shocks does not seem to vary by the selected household attributes (column 4).

4. QUESTIONNAIRE DESIGN'S IMPACT ON POVERTY MEASURED BY PROXIES

Numerous methods to proxy poverty via proxies already exist. We focus on methods that rely on a consumption regression to deduct proxy weights (i.e., beta coefficients), as exemplified by

$$y_i = \beta_j x_{ij} + \varepsilon_i \qquad (2)$$

where y_i is log household consumption expenditures per capita (hereafter referred to as consumption), x_{ij} the vector of proxy variables, and β_j the coefficients (weights) of interest. Examples of such methods include Elbers et al. (2003), Tarozzi (2007), and Mathiassen (2013).

To predict consumption, and in extension thereof poverty and inequality, we utilize the prediction methods developed in Elbers et al. (2003). This prediction method has the advantage of also producing standard errors of poverty and inequality estimates, and implementation is tractable with the PovMap software.[22] To ensure that the results are not model driven and to gauge the sensitivity of poverty and inequality predictions to differences in questionnaire design, we predict consumption with four different models of varying poverty proxy scope. In all cases, we estimate the model in the IHS3 data, using the IHS3 sub-sample interviewed during the months of April-October (i.e., the implementation period for the IHPS), and predict consumption using the IHPS data. To compute predicted poverty rates, we use the official IHS3 poverty line of 37,002 Malawi Kwacha per person per year.

One model that we experiment with is the original WMS prediction model, with updated coefficients from the IHS3 data. In the other three models, variables were selected by *stepwise* in PovMap, which is a statistical method we relied on to avoid selection by researchers. Although the accuracy of the models is not our main interest in comparing predictions based on observationally equivalent samples that are subject to different survey treatment, the complete set of results from the prediction models is provided in the Appendix Tables A6 through A9. The list of possible poverty proxies included in each of the 4 prediction models are as follows:

[22] PovMap software and documentation are freely available on http://iresearch.worldbank.org/PovMap.

1. Experiment only: Only variables derived from the experiment modules that are administered following the household roster;
2. Experiment and non-experiment: Variables derived from the experiment modules as well as demographic, education and locational variables computed from the modules administered prior to the experiment modules;
3. WMS-linked poverty proxies as specified in NSO (2005)[23]; and
4. Non-experiment only: Demographic, education and locational variables computed from the modules administered prior to the experiment modules.

Table 7 presents the differences in the predicted headcount poverty rates and Gini coefficients across different models and sample comparisons.[24] On the whole, variation in questionnaire design is sufficient to generate significant different estimates of both poverty and inequality. Using models 1 through 3, the predicted poverty rate based on the experiment sample is 3 and 7 percentage points *lower* than the predicted poverty rate based on the standard sample (column 1). In all three cases, the predicted poverty rate based on the experiment sample is outside the estimated 95 percent confidence interval for the predicted poverty rate based on standard sample. Similar movements are observed in the predicted Gini coefficients, which are 3 to 4 percentage points *higher* in the short experiment sample (column 1).

The differences in predicted Gini coefficients originating from models 1 through 3 is somewhat expected given the heterogeneity of short questionnaire impact highlighted during the discussion of Table 6. Models 1 through 3 based differences in predicted poverty rates and Gini coefficients also persist for same households as reported in column 2. Working with model 4 (i.e. only with the poverty proxies that are solicited prior to the variation in questionnaire design), there is only 1 percentage point difference in the predicted poverty rate and Gini coefficient between the experiment and standard samples, and the difference is no longer statistically significant. Moreover, looking at column 3, none of the differences between the predictions from the standard interviews of the non-experiment households and the predictions from the standard interviews of the experiment households are statistically significant. Hence, there is strong evidence that the variation in the predicted poverty and inequality statistics is related to the variation in questionnaire design underlying the poverty proxy definitions.

[23] Three variables based on actual expenditures for cooking oil, sugar and soap are not included due to the need to rely on consumer price index series to adjust them over time. In private communication with Astrid Mathiassen, we were able to confirm that the exclusion of these variables from the WMS model does not affect the poverty predictions based on the annual WMS data from 2005 to 2008. We also exclude three binary variables on ownership of bed, iron and refrigerator due to the aforementioned issues in the data collection on durable asset ownership as part of the experiment.

[24] The predicted poverty rates across scenarios are available upon request.

5. CONCLUSION

Our key finding is that observationally equivalent as well as same households answer the same questions differently when interviewed with a short questionnaire vs. the longer counterpart that, in a prior survey round, would have informed the prediction model for a proxy-based poverty measurement exercise. We pick up statistically significant differences in reporting across all topics and types of questions, particularly those related to consumption of non-food and food consumption items, experience of household shocks, subjective welfare and housing. Relying on prediction models based on the national household survey data collected with the standard questionnaire in 2010, we find that the differences in reporting are sufficient to give predicted poverty rates and Gini coefficients that are significantly different from each other. While the difference in predicted poverty estimates ranges from approximately 3 to 7 percentage points depending on the model specification, restricting the proxies to those that are determined prior to the variation in questionnaire design predicts the same poverty rates in both samples.

Although the poverty proxy comparisons are made across different samples without the luxury of the truth, this point matters less in our case precisely because of the focus on proxy-based poverty measurement. The analyst, who would employ the method in the interim years of a complex household survey, also does not know the truth, and would work under the assumption that the available short household survey data would be consistent with the data that would have been collected through the same complex household survey that had generated the poverty prediction model. The short household survey instrument tested in our experiment is one variant out of many that would have been deemed, prior to implementation, sensible and feasible by the research community focused on proxy-based poverty measurement. Abstracting away from possible interview mode effects, our findings should also be of interest to those thinking of using new technologies, such as mobile phones, for collecting consumption or poverty proxy data through succinct interviews.

It is clear that the explanation of the differences is undoubtedly more complex than what is implied by Table 7, and we cannot convincingly map out the mechanisms. The magnitudes of survey treatment effects on questions appearing in the later modules of the standard household questionnaire, such as shocks and subjective well-being, are not larger than those observed earlier as food and non-food consumption related questions. This implies that interview length alone cannot explain the discrepancies. Further, the binary variables are subject to the largest survey treatment effects, and the experiment versions of their respective modules were also the ones where the change in the immediate context of the question was the largest. For instance, in the standard questionnaire, the food consumption module is set up such that a yes/no question is asked for all items to determine consumption in the last 7 days. Once all yes/no questions have been answered they receive follow up questions on quantity consumed, quantity purchased, value of purchases, quantity received as gifts and quantity originating from own-production for items that are reported to be consumed. The experiment version of the same module only includes the yes/no question, asked of only a sub-set of food items. Similar adjustments were made to the modules on non-food consumption and shocks in the context of the experiment. Hence, if standard questionnaire respondents realized the higher likelihood for follow up questions conditional on answering yes to the screening question and intentionally underreported with respect to their counterparts subject to the experiment, this could potentially explain our findings. We do not, however, believe that this possibility applies to our case. As noted above, the experiment was piggybacked onto the second round of a panel survey that used essentially the same standard questionnaire 3 years prior. In addition, the survey treatment effects in Table 5 are not necessarily greater by restricting our analysis to the second half of the fieldwork that included experiment households that had received the standard questionnaire 3 months prior.

Another mechanism at work could be that enumerators may have exerted different levels of effort while administering different questionnaires. One could speculate that with a shorter list of items that are not coupled with follow/up questions, enumerators may have been more dedicated. Since survey treatment effects in Table 4 did not change after including interviewer fixed effects in our regressions, such variation in effort would have to be similar for all interviewers. This variation in effort would also be a source of bias in a typical proxy-based poverty measurement exercise that relies on a different set of interviewers at two different points in time for different questionnaire instruments.

Finally, two broader points relate to direct consumption measurement in household surveys. First, in the case of Malawi, we have shown that the standard questionnaire modules on food and non-food consumption that we seek to proxy take less than 25 minutes to administer as package at the median. Thus, with respect to a household survey for proxy-based poverty measurement, collecting consumption data, in and of itself, may not be as complex and costly as commonly perceived. Here, "perceived" is the operative word as the cost savings in implementing household surveys with a poverty focus net of consumption data is not rigorously documented due to lack of or weaknesses in comparative budgetary and survey process data.

Second, although we do not directly measure consumption in the experiment as well as standard samples, the differences in the propensity to consume food and non-food consumption items suggest that consumption in the standard sample might have been different from consumption in the experiment sample. While we do not have evidence on the relative accuracy of reporting from the experiment and standard samples, under-reporting of consumption is usually assumed to be the main problem in the literature (see, for instance, Beegle et al., 2012). In our case, consumption in the standard sample would appear to be under-reported. Counter examples of systematic over-reporting might exist, though we are unaware of any from general populations in developing countries.

Nevertheless, if there is misreporting in y_i in equation (2) so that $y^{Standard}$ and $y^{Experiment}$ are systematically different from each other, and the same is observed for at least some proxies (x), then β will be biased as well. Following our results in Table 5 and Table 6, it would seem reasonable to assume that the misreporting in x and y are correlated, and have means different from zero. With measurement errors on both sides of the regression, there are no boundaries on direction or size of bias in β (Bound, 2001). Although direct measurement of consumption in household surveys is often considered as the best approximation for true consumption, we can only note that the propensity for reporting consumption is sensitive to questionnaire design, and that consumption regressions from such surveys could be biased due to misreporting.

In future methodological experiments, comparable questionnaire modules could be assigned different orders for different random subsets of the samples that receive experiment versus standard questionnaires, holding the content of the modules, the order of questions in each module, and the interview mode constant. This would, in turn, provide an opportunity to assess whether the reporting differences hold uniformly irrespective of module placement. Similar exercises could be carried out to assess the effect of the order of key questions, holding the content of the modules, the order of modules, and the interview mode constant in alternative questionnaire instruments. These efforts could be complemented by the applications of pretesting techniques, such as cognitive interviews and behavior coding, that could help illuminate cognitive and behavioral processes that play out in answering the same questions as part of different questionnaires (Presser et al., 2004). Moving forward, household survey operations designed for proxy-based poverty measurement should, prior to full roll-out, consider piloting their instruments in parallel with the questionnaire instruments from which they have evolved. This methodological exercise could be designed as a randomized household survey experiment to test whether the data for poverty predictors differ depending on whether they were solicited in an experiment versus a standard questionnaire.

REFERENCES

Andrews, F. M. (1984). Construct validity & error components of survey measures: a structural modelling approach. *Public Opinion Quarterly*, 48, pp. 409–442.

Beegle, K., De Weerdt, J., Friedman, J., & Gibson, J. (2012). Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*, 98.1, pp. 3–18.

Bound, J., Brown, C., & Mathiowetz, N. (2001). In J. Heckman & E. Leamer (Eds.), *Measurement error in survey data* (3707–3843), Elsevier: Amsterdam.

Christiaensen, L., Lanjouw, P. Luoto, J., & Stifel, D. (2012). Small area estimation-based prediction methods to track poverty: validation & applications. *Journal of Economic Inequality*, 10.2, pp. 267–297.

Douidich, M., Ezzrari, A., Van der Weide, R., & Verme, P. (2013). Estimating quarterly poverty rates using labor force surveys: a primer, World Policy Research Working Paper No. 6466.

Eckman, S., Kreuter, F., Kirchner, A., Jaeckle, A., Tourangeau, R., & Presser, S. (Forthcoming). Assessing the mechanisms of misreporting to filter questions in surveys. *Public Opinion Quarterly.*

Elbers, C., Lanjouw, J. O., & Lanjouw. P. (2003). Micro-level estimation of poverty & inequality. *Econometrica*, 71.1, pp. 355–364.

Herzog, R. A., & Bachman, G. J. (1981). Effects of questionnaire length on response quality. *The Public Quarterly,* 45.4, pp. 549–559.

Hess, J., Moore, J., Pascale, J., Rothgeb, J., & Keeley, C. The effects of person-level versus household-level questionnaire design on survey estimates & data quality. *The Public Opinion Quarterly*, 65.4, pp. 574–584.

Houssou, N., & Zeller, M. (2011). To target or not to target? The costs, benefits, & impacts of indicator-based targeting. *Food Policy*, 36.5, pp. 627–637.

Johnson, W. R., Sieveking, N. A., & Clanton III, E. S. (1974). Effects of alternative positioning of open-ended questions in multiple-choice questionnaires. *Journal of Applied Psychology,* 59.6, pp. 776–778.

Kraut, A. I., Wolfson, A. D., & Rothenberg, A. (1975). Some effects of position on opinion survey items. *Journal of Applied Psychology,* 60, pp. 774–776.

Kreuter, F., McCulloch, S., Presser, S., & Tourangeau, R. (2011). The effects of asking filter questions in interleafed versus grouped format. *Sociological Methods & Research*, 40.1, pp. 88–104.

Mathiassen, A. (2013). Testing prediction performance of poverty models: empirical evidence from Uganda. Review of Income & Wealth, 59.1, pp. 91–112.

National Statistical Office (NSO) (2014). Malawi integrated household panel survey 2013 basic information document. Retrieved from http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1233781970982/5800988-1271185595871/6964312-1404828635943/IHPS_BID_FINAL.pdf

Newhouse, D., Shivakumaran, S., Takamatsu, S., & Yoshida, N. (2014). How survey-to-survey imputation can fail. World Bank Policy Research Working Paper Series No. 6961.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., and Singer, E. (2004). Methods of testing and evaluating survey questions. *Public Opinion Quarterly*, 68.1, pp. 109–130.

Tarozzi, A. (2007). Calculating comparable statistics from incomparable surveys, with an application to poverty in India. *Journal of Business & Economic Statistics*, 25.3, pp. 314–336.

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000): *The psychology of survey response*. Cambridge University Press: New York, NY & Cambridge, U.K.

Vu, L., & Baulch, B. (2011). Assessing alternative poverty proxy methods in rural Vietnam. *Oxford Development Studies*, 39.3, pp. 339–367.

Table 1: Sample Split by Visit and Interview Type

	Visit 1	Visit 2	Total
Standard Interviews (S)	1,428 *Sample A*	1,394 *Sample B*	2,822
Short Interviews (L)	1,394 *Sample B*	1,428 *Sample A*	2,822
Experiment Sub-Sample Out of Short Interviews (E)	372	393	765
Implied Records in Analysis (S+E)	1,800	1,787	3,587

Table 2: Module & Interview Durations

	Standard Interview		*Experiment Interview*	
Module	*Module Duration*	*Time Elapsed Prior to Module*	*Module Duration*	*Time Elapsed Prior to Module*
Household Roster	9	--	9	--
Education, Health & Labor	24	9	--	--
Housing	6	34	2	9
Food Consumption	16	40	2	11
Food Security	2	58	--	--
Non-Food Consumption (All 3 Modules)	9	61	2	13
Durable Goods	3	71	2	16
Farm Implements, Machinery, and Structures	3	75	--	--
Household Enterprises	2	79	--	--
Children Living Elsewhere	1	83	--	--
Other Income	2	86	--	--
Gifts Given Out	1	88	--	--
Social Safety Nets	2	90	--	--
Credit	2	93	--	--
Subjective Assessment of Well-Being	5	96	2	21
Shocks and Coping Strategies	5	101	3	18
Child Anthropometry	1	106	--	--
Filter for Agriculture & Fishery Questionnaires	1	108	--	--
Total Interview Duration	**109**		**23**	

Note: Median durations are reported in minutes. Education, Health & Labor were separate modules but were not time-stamped separately - there were time stamps only at the beginning of the Education module and at the end of the Labor module.

Table 3: Module-Specific Breakdown of Poverty Proxies Subject to Statistically Significant Survey Treatment Effects

	Total # of Poverty Proxies	*# of Poverty Proxies Subject to Statistically Significant Survey Treatment Effects*	
		Experiment vs. Standard	*Same Households: Experiment vs. Standard*
	(1)	(2)	(3)
Binary			
Food	22	7	8
Non-Food	25	14	13
Shocks	23	6	6
Ordered Categorical			
Housing	4	1	1
Subjective Welfare	4	2	2
Durable Assets	4	2	2
Continuous			
Cell Phone Expenditures	1	1	0
TOTAL	83	33	32

Note: Binary, ordered categorical, and continuous variable related differences in reporting in columns 2 and 3 are based on multivariate Logit, Ordered Logit, and Ordinary Least Squares regressions, respectively, specified in accordance with Equation (1). The regressions are weighted and take into account clustering and stratification; The statistical significance level used is 10 percent.

Table 4: Selected Regressions Results Based on Pooled Data

	Pooled Binary Poverty Proxies					
	Experiment vs. Standard			*Same Households: Experiment vs. Standard*		
	(1)	(2)	(3)	(4)	(5)	(6)
Model	1	2	3	1	2	3
Controls	NO	YES	YES	NO	YES	YES
Interviewer Fixed Effects	NO	NO	YES	NO	NO	YES
Experiment	0.027***	0.023***	0.022***	0.027***	0.023***	0.022***
	(0.005)	(0.005)	(0.005)	(0.004)	(0.004)	(0.004)
Observations	197,513	197,443	197,443	107,080	107,010	107,010
	Pooled Ordered Categorical Poverty Proxies (3 Categories)					
Experiment	0.015*	0.020**	0.021***	0.014**	0.018**	0.016**
	(0.009)	(0.008)	(0.008)	(0.007)	(0.007)	(0.007)
Observations	8,465	8,462	8,462	4,590	4,587	4,587
	Pooled Ordered Categorical Poverty Proxies (4 Categories)					
Experiment	-0.009***	-0.006*	-0.006*	-0.003	-0.003	-0.003
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Observations	8,465	8,462	8,462	4,589	4,586	4,586
	Pooled Ordered Categorical Poverty Proxies (6 Categories)					
Experiment	0.031***	0.032***	0.031***	0.025**	0.023**	0.023**
	(0.011)	(0.011)	(0.012)	(0.010)	(0.011)	(0.011)
Observations	8,466	8,463	8,463	4,590	4,587	4,587
	Pooled Ordered Categorical Poverty Proxies (11 Categories)					
Experiment	0.002	0.002	0.002	-0.004	-0.004	-0.004
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)
Observations	2,798	2,797	2,797	1,514	1,513	1,513

Note: Experiment is equal to 1 if the household was subject to the experiment questionnaire treatment, and 0 otherwise. The estimations use the pooled (binary or ordered categorical) data at the household level; The results for the pooled binary and pooled ordered categorical poverty proxies originate from multivariate Logit and Ordered Logit regressions, respectively. While multivariate Logit regression results are marginal effects, the multivariate Ordered Logit regression results represent marginal effects on the probability of being in the lowest category. The control variables, as specified in Equation 1, are included when noted. The regressions are weighted and take into account clustering and stratification; The results are robust to varying the set of control variables. ***/**/* indicate statistical significance at the 1/5/10 percent level.

Table 5: Heterogeneity of Experiment Questionnaire Treatment Impact on Pooled Binary Poverty Proxies Across Sample Comparisons & Questionnaire Modules

	All		**Food**		**Non-Food**		**Shocks**	
Sample Comparison	*Experiment vs. Standard*	*Same Households: Experiment vs. Standard*	*Experiment vs. Standard*	*Same Households: Experiment vs. Standard*	*Experiment vs. Standard*	*Same Households: Experiment vs. Standard*	*Experiment vs. Standard*	*Same Households: Experiment vs. Standard*
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Overall	0.023***	0.023***	0.026***	0.031***	0.029***	0.032***	0.014**	0.006
	(0.005)	(0.004)	(0.008)	(0.007)	(0.006)	(0.006)	(0.006)	(0.004)
Observations	197,513	107,080	62,082	33,658	70,529	38,235	64,902	35,187
1st Half of Fieldwork	0.023***		0.022*		0.035***		0.010	
	(0.008)		(0.013)		(0.010)		(0.010)	
Observations	98,466		30,952		35,156		32,358	
2nd Half of Fieldwork	0.022***		0.025**		0.023**		0.019**	
	(0.007)		(0.012)		(0.009)		(0.008)	
Observations	99,047		31,130		35,373		32,544	

Note: The reported coefficients and standard errors are those associated with the binary variable identifying whether a household was subject to the experiment questionnaire treatment. The estimations are based on Logit regressions, using the pooled data at the household level for all 70 binary poverty proxies from food, non-food and shocks modules. The control variables, as specified in Equation 1, are included but not reported. The regressions are weighted and take into account clustering and stratification. The results are robust to varying the set of control variables and/or including interviewer fixed effects. ***/**/* indicate statistical significance at the 1/5/10 percent level.

Table 6: Heterogeneity of Experiment Questionnaire Treatment Impact on Pooled Binary Poverty Proxies by Selected Household Attributes

	Sample Comparison: Experiment vs. Standard			
	All	**Food**	**Non-Food**	**Shocks**
	(1)	(2)	(3)	(4)
Experiment*Female Head †	-0.011	-0.009	-0.017	-0.007
	(0.011)	(0.020)	(0.014)	(0.011)
Experiment*Head Age (Years)	0.001**	0.001***	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Experiment*Highest HH Education: No Education †	0.020**	0.011	0.022	0.022**
	(0.010)	(0.019)	(0.015)	(0.011)
Experiment*Highest HH Education: Primary †	-0.000	-0.031	-0.012	0.039***
	(0.012)	(0.023)	(0.020)	(0.014)
Experiment*Household Size	0.010***	0.012**	0.020***	0.001
	(0.003)	(0.005)	(0.005)	(0.004)
Experiment*# of Dependents	-0.009*	-0.009	-0.019***	-0.000
	(0.005)	(0.008)	(0.007)	(0.005)
Experiment*Rural †	-0.022**	-0.052***	-0.032**	0.004
	(0.010)	(0.017)	(0.016)	(0.016)
Observations	197,443	62,060	70,504	64,879

Note: † indicates a binary variable. The reported coefficients and standard errors are marginal effects associated with the interactions between the selected household attributes and the binary variable identifying whether a household was subject to the experiment questionnaire treatment. The estimations are based on Logit regressions, using the pooled data at the household level for all 70 binary poverty proxies from food, non-food and shocks modules. The control variables, as specified in Equation 1, are included but not reported. The regressions are weighted and take into account clustering and stratification. The results are robust to varying the set of control variables and/or including interviewer fixed effects. ***/**/* indicate statistical significance at the 1/5/10 percent level.

Table 7: Differences in Predictions for Headcount Poverty Rate and Gini Coefficient across Prediction Models & Sample Comparisons

	Differences in Headcount Poverty Rate Predictions		
	Prediction from Standard Interviews - Prediction from Experiment Interviews	*Same Households: Prediction from Standard Interviews - Prediction from Experiment Interviews*	*Prediction from Standard Interviews of Non-Experiment Households - Prediction from Standard Interviews of Experiment Households*
Model	(1)	(2)	(3)
1. Experiment Only	**0.05**	**0.03**	0.01
2. Experiment & Non-Experiment	**0.07**	**0.06**	0.00
3. WMS Model	**0.03**	**0.04**	0.00
4. Non-Experiment Only	0.01	0.00	0.00
	Differences in Gini Coefficient Predictions		
	Prediction from Standard Interviews - Prediction from Experiment Interviews	*Same Households: Prediction from Standard Interviews - Prediction from Experiment Interviews*	*Prediction from Standard Interviews of Non-Experiment Households - Prediction from Standard Interviews of Experiment Households*
Model	(1)	(2)	(3)
1. Experiment Only	**-0.03**	**-0.02**	-0.01
2. Experiment & Non-Experiment	**-0.04**	**-0.03**	-0.01
3. WMS Model	**-0.03**	**-0.02**	-0.01
4. Non-Experiment Only	-0.01	0.00	-0.01

Note: Bold indicates scenarios in which the experiment sample based prediction is outside of the 95 percent confidence interval for the prediction based on the comparator sample (standard interviews for columns 1 and 2, standard interviews of non-experiment households in column 3).

APPENDIX

Table A1: Sample Means by Household Experiment Status

	Non-Experiment	*Experiment*
Household Size	5.02	5.03
# of HH Members: Age 0-5	0.91	0.89
# of HH Members: Age 6-14	1.45	1.46
# of HH Members: Female, Age 15-39	0.93	0.94
# of HH Members: Male, Age 15-39	0.90	0.86
# of HH Members: Female, Age 40-59	0.28	0.30
# of HH Members: Male, Age 40-59	0.26	0.28
# of HH Members: Age 60+	0.28	0.30
Number of Baseline Individuals	4.15	4.26
Head of Household Attributes		
Age (Years)	44.63	45.58
Female	0.26	0.28
Ethnicity		
Chewa †	0.61	0.61
Tumbuka †	0.05	0.06
Other †	0.33	0.33
Highest Education		
None †	0.78	0.77
Primary †	0.09	0.10
Junior High †	0.07	0.06
Secondary & Above †	0.06	0.07
Religion		
Christianity †	0.77	0.77
Islam †	0.17	0.16
Other†	0.06	0.07
Marital Status		
Union, Monogamous †	0.68	0.67
Union, Polygamous †	0.07	0.06
Separated †	0.05	0.06
Divorced †	0.06	0.05
Widowed/Widower †	0.14	0.14
Never Married †	0.01	0.02
Household Highest Education		
None †	0.64	0.63
Primary †	0.14	0.14
Junior High †	0.12	0.12
Secondary & Above †	0.10	0.11
Household Location		
Rural	0.87	0.85
Northern Region	0.09	0.09
Central Region	0.44	0.45
Southern Region	0.47	0.45
Distance to Baseline Location (KMs)	1.31	1.29
Observations	2,057	765

Note: † indicates a binary variable. No mean comparison is statistically significant at least at the 10 percent level.

Table A2: Marginal Effects of Experiment Questionnaire Treatment Across Models, Sample Specifications & Categories of Pooled Ordered Categorical Poverty Proxies (3 Categories)

	Model 1 - Experiment vs. Standard		
	(1)	(2)	(3)
Experiment	0.015*	0.000	-0.016*
	(0.009)	(0.001)	(0.009)
	Model 2 - Experiment vs. Standard		
	(1)	(2)	(3)
Experiment	0.020**	0.001	-0.020**
	(0.008)	(0.001)	(0.009)
	Model 3 - Experiment vs. Standard		
	(1)	(2)	(3)
Experiment	0.021***	0.000	-0.022***
	(0.008)	(0.001)	(0.008)
	Model 1 - Same Households: Experiment vs. Standard		
	(1)	(2)	(3)
Experiment	0.014**	-0.000	-0.014**
	(0.007)	(0.001)	(0.007)
	Model 2 - Same Households: Experiment vs. Standard		
	(1)	(2)	(3)
Experiment	0.018**	-0.000	-0.017**
	(0.007)	(0.001)	(0.007)
	Model 3 - Same Households: Experiment vs. Standard		
	(1)	(2)	(3)
Experiment	0.016**	-0.000	-0.015**
	(0.007)	(0.001)	(0.006)

Note: Experiment is equal to 1 if the household was subject to the experiment questionnaire treatment, and 0 otherwise. Models are as defined in Table 4. The estimations use the ordered categorical data with 3 categories, pooled at the household level. The results originate from Ordered Logit regressions. The results are the marginal effects on the probability of being in each category. The regressions are weighted and take into account clustering and stratification. ***/**/* indicate statistical significance at the 1/5/10 percent level.

Table A3: Marginal Effects of Experiment Questionnaire Treatment Across Models, Sample Specifications & Categories of Pooled Ordered Categorical Poverty Proxies (4 Categories)

	Model 1 - Experiment vs. Standard			
	(1)	(2)	(3)	(4)
Experiment	-0.009***	0.001*	0.019***	0.005***
	(0.003)	(0.001)	(0.007)	(0.002)
	Model 2 - Experiment vs. Standard			
	(1)	(2)	(3)	(4)
Experiment	-0.006*	0.001	0.012*	0.003*
	(0.003)	(0.000)	(0.006)	(0.001)
	Model 3 - Experiment vs. Standard			
	(1)	(2)	(3)	(4)
Experiment	-0.006*	0.001	0.011*	0.003*
	(0.003)	(0.000)	(0.006)	(0.001)
	Model 1 - Same Households: Experiment vs. Standard			
	(1)	(2)	(3)	(4)
Experiment	-0.003	0.000	0.006	0.002
	(0.002)	(0.000)	(0.004)	(0.001)
	Model 2 - Same Households: Experiment vs. Standard			
	(1)	(2)	(3)	(4)
Experiment	-0.003	0.000	0.006	0.001
	(0.002)	(0.000)	(0.005)	(0.001)
	Model 3 - Same Households: Experiment vs. Standard			
	(1)	(2)	(3)	(4)
Experiment	-0.003	0.000	0.005	0.001
	(0.002)	(0.000)	(0.005)	(0.001)

Note: Experiment is equal to 1 if the household was subject to the experiment questionnaire treatment, and 0 otherwise. Models are as defined in Table 4. The estimations use the ordered categorical data with 4 categories, pooled at the household level. The results originate from Ordered Logit regressions. The results are the marginal effects on the probability of being in each category. The regressions are weighted and take into account clustering and stratification. ***/**/* indicate statistical significance at the 1/5/10 percent level.

Table A4: Marginal Effects of Experiment Questionnaire Treatment Across Models, Sample Specifications & Categories of Pooled Ordered Categorical Poverty Proxies (6 Categories)

	Model 1 - Experiment vs. Standard					
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment	0.031***	0.010**	-0.023***	-0.012***	-0.004***	-0.002**
	(0.011)	(0.004)	(0.008)	(0.004)	(0.002)	(0.001)
	Model 2 - Experiment vs. Standard					
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment	0.032***	0.012***	-0.026***	-0.012***	-0.004***	-0.001**
	(0.011)	(0.004)	(0.009)	(0.004)	(0.002)	(0.001)
	Model 3 - Experiment vs. Standard					
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment	0.031***	0.013**	-0.028***	-0.011***	-0.004**	-0.001**
	(0.012)	(0.005)	(0.010)	(0.004)	(0.001)	(0.001)
	Model 1 - Same Households: Experiment vs. Standard					
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment	0.025**	0.007**	-0.019**	-0.009**	-0.003**	-0.002**
	(0.010)	(0.003)	(0.007)	(0.004)	(0.001)	(0.001)
	Model 2 - Same Households: Experiment vs. Standard					
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment	0.023**	0.007**	-0.019**	-0.008**	-0.002**	-0.001*
	(0.011)	(0.003)	(0.009)	(0.004)	(0.001)	(0.001)
	Model 3 - Same Households: Experiment vs. Standard					
	(1)	(2)	(3)	(4)	(5)	(6)
Experiment	0.023**	0.008**	-0.021**	-0.007**	-0.002**	-0.001*
	(0.011)	(0.004)	(0.010)	(0.003)	(0.001)	(0.001)

Note: Experiment is equal to 1 if the household was subject to the experiment questionnaire treatment, and 0 otherwise. Models are as defined in Table 4. The estimations use the ordered categorical data with 6 categories, pooled at the household level. The results originate from Ordered Logit regressions. The results are the marginal effects on the probability of being in each category. The regressions are weighted and take into account clustering and stratification. ***/**/* indicate statistical significance at the 1/5/10 percent level.

Table A5: Marginal Effects of Experiment Questionnaire Treatment Across Models, Sample Specifications & Categories of Pooled Ordered Categorical Poverty Proxies (11 Categories)

	Model 1 - Experiment vs. Standard										
	Categories										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Experiment	0.002	0.004	0.005	0.002	-0.001	-0.002	-0.002	-0.002	-0.001	-0.002	-0.004
	(0.003)	(0.008)	(0.008)	(0.003)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.006)
	Model 2 - Experiment vs. Standard										
	Categories										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Experiment	0.002	0.004	0.005	0.002	-0.001	-0.002	-0.002	-0.002	-0.001	-0.002	-0.004
	(0.003)	(0.008)	(0.008)	(0.003)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.006)
	Model 3 - Experiment vs. Standard										
	Categories										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Experiment	0.002	0.004	0.005	0.002	-0.001	-0.002	-0.002	-0.002	-0.001	-0.002	-0.004
	(0.003)	(0.008)	(0.008)	(0.003)	(0.001)	(0.003)	(0.003)	(0.003)	(0.002)	(0.004)	(0.006)
	Model 1 - Same Households: Experiment vs. Standard										
	Categories										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Experiment	-0.004	-0.010	-0.009	-0.002	0.002	0.004	0.003	0.003	0.002	0.003	0.008
	(0.002)	(0.006)	(0.006)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.005)
	Model 2 - Same Households: Experiment vs. Standard										
	Categories										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Experiment	-0.004	-0.010	-0.009	-0.002	0.002	0.004	0.003	0.003	0.002	0.003	0.008
	(0.002)	(0.006)	(0.006)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.005)
	Model 3 - Same Households: Experiment vs. Standard										
	Categories										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Experiment	-0.004	-0.010	-0.009	-0.002	0.002	0.004	0.003	0.003	0.002	0.003	0.008
	(0.002)	(0.006)	(0.006)	(0.002)	(0.001)	(0.002)	(0.002)	(0.002)	(0.001)	(0.002)	(0.005)

Note: Experiment is equal to 1 if the household was subject to the experiment questionnaire treatment, and 0 otherwise. Models are as defined in Table 4. The estimations use the ordered categorical data with 11 categories, pooled at the household level. The results originate from Ordered Logit regressions. The results are the marginal effects on the probability of being in each category. The regressions are weighted and take into account clustering and stratification. ***/**/* indicate statistical significance at the 1/5/10 percent level.

Table A6: Prediction Model 1- Experiment Only

	Coefficient	Standard Error	P-value
Intercept	10.02	0.04	0.00
Food Consumption			
Bread †	0.13	0.02	0.00
Groundnuts †	0.15	0.01	0.00
Brown Beans †	0.04	0.01	0.00
Eggs †	0.17	0.02	0.00
Cassava †	0.05	0.01	0.00
Maize - Fine Flour †	0.13	0.01	0.00
Meat †	0.21	0.01	0.00
Milk †	0.14	0.02	0.00
Rice †	0.18	0.02	0.00
Nkhwani †	0.07	0.01	0.00
Tomato †	0.08	0.02	0.00
Oil †	0.08	0.02	0.00
Sugar †	0.12	0.02	0.00
Chips †	0.08	0.02	0.00
Non-Food Consumption			
Public Transport - Bus/Minibus †	0.15	0.02	0.00
Men's Jackets †	0.18	0.05	0.00
Men's Clothing (Any Type) †	0.08	0.02	0.00
Men's Shirts †	0.09	0.03	0.00
Candles †	0.08	0.02	0.00
Cigarettes/Tobacco †	0.14	0.03	0.00
Other Personal Cosmetic Products †	0.09	0.02	0.00
Toothpaste/Toothbrush †	0.09	0.02	0.00
Boy's Shoes †	-0.14	0.03	0.00
Shoes (Any Type) †	0.09	0.02	0.00
Bar Soap †	0.08	0.03	0.00
Girl's Shoes †	-0.12	0.02	0.00
Clothes Soap (Powder) †	0.06	0.02	0.00
Newspapers/Magazines †	0.40	0.04	0.00
Shocks			
Unusually High Level of Livestock Disease †	0.08	0.03	0.00
Birth in the Household †	-0.23	0.04	0.00
Drought †	-0.04	0.01	0.01
Earthquake †	-0.10	0.03	0.00
Unusually High Prices for Food †	-0.07	0.02	0.00
Theft †	0.15	0.03	0.00
Housing Characteristics			
Dwelling Owned †	-0.08	0.02	0.00
Floor: Sand/Mud †	-0.18	0.02	0.00
Durable Assets			
Household Head Sleeps Under Blanket & Sheets †	0.09	0.02	0.00
Household Head Number of Changes of Clothes 8+ †	0.09	0.02	0.00
Observations		6502	
R2		0.56	
Adjusted R2		0.56	

Note: † indicates a binary variable.

Table A7: Prediction Model 2 - Experiment & Non-Experiment

	Coefficient	Standard Error	P-value
Intercept	10.60	0.04	0.00
Demographics & Education			
Household Size	-0.15	0.01	0.00
Age of Household Head (Years)	0.00	0.00	0.00
Dependency Ratio	-0.54	0.04	0.00
Highest Educational Qualification in Household †	0.06	0.01	0.00
Number of Household Members 60+	0.08	0.02	0.00
Number of Household Members 0-14	0.07	0.01	0.00
Food Consumption			
Bread †	0.11	0.02	0.00
Chips †	0.07	0.02	0.00
Sugar †	0.12	0.01	0.00
Sweet Potatoes †	0.05	0.01	0.00
Eggs †	0.14	0.01	0.00
Groundnuts †	0.11	0.01	0.00
Meat †	0.20	0.01	0.00
Brown Beans †	0.08	0.01	0.00
Milk †	0.15	0.02	0.00
Nkhwani †	0.06	0.01	0.00
Tomatoes †	0.06	0.01	0.00
Oil †	0.09	0.01	0.00
Maize - Fine Flour †	0.10	0.01	0.00
Rice †	0.12	0.01	0.00
Non-Food Consumption			
Public Transportation: Bus/Minibus †	0.15	0.01	0.00
Bar Soap †	0.12	0.02	0.00
Clothes Soap (Powder) †	0.10	0.01	0.00
Candles †	0.06	0.02	0.00
Charcoal †	0.10	0.02	0.00
Men's Jackets †	0.14	0.03	0.00
Men's Trousers †	0.04	0.02	0.01
Toothpaste/Toothbrush †	0.07	0.01	0.00
Public Transportation: Other †	0.19	0.05	0.00
Men's Shoes †	0.07	0.02	0.00
Men's Shirts †	0.06	0.02	0.00
Cigarettes/Tobacco †	0.14	0.02	0.00
Newspapers/Magazines †	0.34	0.03	0.00
Other Personal Cosmetic Products †	0.10	0.01	0.00
Shocks			
Unusually High Level of Livestock Disease †	0.08	0.02	0.00
Unusually High Prices for Food †	-0.05	0.01	0.00
Unusually High Prices for Agricultural Output †	0.05	0.02	0.00
Death of a Non-Income Earning Household Member †	-0.09	0.03	0.00
Theft †	0.09	0.02	0.00
Housing Characteristics			
Roof: Grass †	-0.07	0.01	0.00
Floor: Sand/Mud †	-0.11	0.02	0.00
Dwelling Owned †	0.04	0.01	0.00

Table A7 (Cont'd)

	Coefficient	Standard Error	P-value
People Per Room	-0.07	0.00	0.00
Durable Assets	0.00	0.00	0.00
Cell Phone †	0.06	0.01	0.00
Household Head Number of Changes of Clothes 0-2 †	-0.04	0.01	0.01
Radio †	0.04	0.01	0.00
Household Head Sleeps Under Blanket & Sheets †	0.06	0.01	0.00
Location Fixed Effects			
District 101 †	-0.29	0.04	0.00
District 102 †	-0.19	0.03	0.00
District 105 †	-0.15	0.02	0.00
District 201 †	0.17	0.03	0.00
District 202 †	0.19	0.03	0.00
District 203 †	0.31	0.04	0.00
District 204 †	0.14	0.02	0.00
District 205 †	0.09	0.03	0.00
District 208 †	-0.07	0.02	0.01
District 310 †	-0.29	0.03	0.00
District 311 †	-0.33	0.03	0.00
District 312 †	-0.11	0.03	0.00
Rural †	-0.11	0.02	0.00
Observations		6502	
R2		0.77	
Adjusted R2		0.76	

Note: † indicates a binary variable.

Table A8: Prediction Model 3 - WMS Model

	Coefficient	Standard Error	P-value
Intercept	10.81	0.03	0.00
Demographics & Education			
Age of Household Head (Years)	0.01	0.00	0.00
Household Size	-0.14	0.01	0.00
Highest Educational Qualification in Household †	0.07	0.01	0.00
Dependency Ratio	-0.47	0.04	0.00
Number of Household Members 0-14	0.05	0.01	0.00
Food Consumption			
Bread †	0.16	0.02	0.00
Eggs †	0.20	0.01	0.00
Meat †	0.22	0.01	0.00
Milk †	0.18	0.02	0.00
Oil †	0.13	0.01	0.00
Rice †	0.14	0.01	0.00
Sugar †	0.16	0.01	0.00
Non-Food Consumption			
Men's Other Clothing †	0.10	0.04	0.01
Shoes (Any Type) †	0.14	0.01	0.00
Toothpaste/Toothbrush †	0.14	0.01	0.00
Public Transportation: Other †	0.25	0.06	0.00
Housing Characteristics			
Roof: Grass †	-0.04	0.02	0.02
Floor: Sand/Mud †	-0.14	0.02	0.00
People Per Room	-0.06	0.01	0.00
Durable Assets			
Household Head Number of Changes of Clothes	0.00	0.00	0.00
Radio †	0.05	0.01	0.00
Cell Phone †	0.08	0.02	0.00
Household Head Sleeps Under Blanket & Sheets †	0.07	0.01	0.00
Observations		6502	
R2		0.68	
Adjusted R2		0.68	

Note: † indicates a binary variable.

Table A9: Prediction Model 4 - Non-Experiment Only

	Coefficient	Standard Error	P-value
Intercept	10.98	0.04	0.00
Demographics & Education			
Age of Household Head (Years)	0.00	0.00	0.00
Dependency Ratio	-0.70	0.06	0.00
Household Size	-0.15	0.01	0.00
Highest Educational Qualification in Household †	0.24	0.01	0.00
Number of Household Members 60+	0.11	0.02	0.00
Number of Household Members 0-14	0.06	0.01	0.00
Location Fixed Effects			
District 101 †	-0.26	0.06	0.00
District 103 †	0.35	0.07	0.00
District 104 †	0.26	0.06	0.00
District 201 †	0.34	0.04	0.00
District 202 †	0.40	0.04	0.00
District 203 †	0.42	0.06	0.00
District 204 †	0.26	0.04	0.00
District 205 †	0.24	0.04	0.00
District 206 †	0.13	0.03	0.00
District 207 †	0.10	0.04	0.01
District 209 †	0.13	0.04	0.00
District 210 †	0.12	0.04	0.00
District 302 †	-0.07	0.04	0.05
District 303 †	0.22	0.04	0.00
District 304 †	0.30	0.05	0.00
District 305 †	0.33	0.04	0.00
District 307 †	0.25	0.03	0.00
District 310 †	-0.50	0.04	0.00
District 311 †	-0.51	0.05	0.00
District 315 †	0.22	0.04	0.00
Rural †	0.28	0.03	0.00
Obs		6502	
R2		0.48	
Adjusted R2		0.48	

Note: † indicates a binary variable.

EXPERIMENT QUESTIONNAIRE MODULES

MODULE B1: HOUSING

ENUMERATOR: RECORD START TIME FOR MODULE B1: HOURS MINUTES

ENUMERATOR: RECORD PRIMARY RESPONDENT ID FOR MODULE B1: ID

ENUMERATOR: RECORD END TIME FOR MODULE B1: HOURS MINUTES

B101	B102	B103	B104	B105	B106	B107	B108
Do you own or are purchasing this house, is it provided to you by an employer, do you use it for free, or do you rent this house?	THE ROOF OF THE MAIN DWELLING IS PREDOMINANTLY MADE OF WHAT MATERIAL?	THE FLOOR OF THE MAIN DWELLING IS PREDOMI-NANTLY MADE OF WHAT MATERIAL?	How many separate rooms do the members of your household occupy? (DO NOT COUNT BATHROOMS, TOILETS, STOREROOMS, OR GARAGE)	How many working cell phones in total does your household own? IF NONE, RECORD 0.	Estimate the total cost for all cell phone service for all household members last month?	What kind of toilet facility does your household use?	Do any members of your household sleep under a bed net to protect against mosquitoes at some time during the year?
OWNED. . . . 1 BEING PURCHASED . 2 EMPLOYER PROVIDES. . ..3 FREE, AUTHORIZED . .4 FREE, NOT AUTHORIZED. ..5 RENTED. 6	GRASS......1 IRON SHEETS.....2 CLAY TILES......3 CONCRETE...4 PLASTIC SHEETING...5 OTHER (SPECIFY)..6	SAND. ....1 SMOOTHED MUD. .....2 SMOOTH CEMENT....3 WOOD. .. .4 TILE.. . .5 OTHER (SPECIFY).6	NUMBER	NUMBER	MK	FLUSH TOILET. . 1 VIP LATRINE. .2 TRADIT. LATRINE W/ROOF. . 3 TRADIT. LATRINE W/O ROOF. 4 NONE. . . .5 OTHER (SPECIFY) 6	YES..1 NO...2

MODULE B2: FOOD CONSUMPTION

ENUMERATOR: RECORD START TIME FOR MODULE B2: HOURS MINUTES

ENUMERATOR: RECORD PRIMARY RESPONDENT ID FOR MODULE B2: ID

ENUMERATOR: RECORD END TIME FOR MODULE B2: HOURS MINUTES

Over the past one week (7 days), did you or others in your household consume any [. . .]? INCLUDE FOOD BOTH EATEN COMMUNALLY IN THE HOUSEHOLD AND THAT EATEN SEPARATELY BY INDIVIDUAL HOUSEHOLD MEMBERS.	B201 YES.1 NO..2> > NEXT ITEM	B202 ITEM CODE	ITEM	B201 YES.1 NO..2>> NEXT ITEM	B202 ITEM CODE
Maize *ufa mgaiwa* (normal flour) *		101	Eggs		501
Maize *ufa* refined (fine flour) *		102	Beef		504
Maize *ufa madeya* (bran flour) *		103	Goat		505
Rice		106	Pork		506
Bread		111	Chicken		508
Cassava tubers*		201	Other poultry-guinea fowl,doves,etc.		509
White sweet potato*		203	Fresh milk		701
Bean, brown*		302	Sugar		801
Groundnut*		304	Cooking oil		803
Nkhwani*		404	Chips (vendor)		821
Tomato*		408	Mandazi, doughnut (vendor)		827

MODULES B3, B4 & B5: NON-FOOD EXPENDITURES

MODULE B3: ONE WEEK RECALL

Over the past one week (7 days), did your household purchase or pay for any [...]?	B301 YES.1 NO..2>> NEXT ITEM	B302 ITEM CODE
Charcoal		101
Paraffin or kerosene		102
Cigarettes or other tobacco		103
Candles		104
Matches		105
Newspapers or magazines		106
Public transport - Bicycle Taxi		107
Public transport - Bus/Minibus		108
Public transport - Other (Truck, Oxcart, Etc..)		109

MODULE B4: ONE MONTH RECALL

Over the past one month, did your household purchase or pay for any [...]?	B401 YES.1 NO..2>> NEXT ITEM	B402 ITEM CODE
Bar soap (body soap or clothes soap)		202
Clothes soap (powder)		203
Toothpaste, toothbrush		204
Glycerine, Vaseline, skin creams		206
Other personal products (shampoo, razor blades, cosmetics, hair products, etc.)		207
Batteries		220
Recharging batteries, cell phones		221

MODULE B5: THREE MONTH RECALL

Over the past three months, did your household purchase or pay for any [...]?	B501 YES.1 NO..2>> NEXT ITEM	B502 ITEM CODE
Men's trousers		308
Men's shirts		309
Men's jackets		310
Men's undergarments		311
Men's other clothing		312
Boy's shoes		322
Men's shoes		323
Girl's shoes		324
Lady's shoes		325

ENUMERATOR: RECORD START TIME FOR MODULE B3: HOURS MINUTES

ENUMERATOR: RECORD PRIMARY RESPONDENT ID FOR MODULE B3,B4,B5: ID

ENUMERATOR: RECORD END TIME FOR MODULE B5: HOURS MINUTES

MODULE B6: ASSETS

ENUMERATOR: RECORD START TIME FOR MODULE B6: ____ HOURS ____ MINUTES

ITEM	B601 How many [ITEM]s do you own? IF NONE, RECORD ZERO. NUMBER	B602 ITEM CODE
Mortar/pestle (*mtondo*)		501
Bed		502
Table		503
Chair		504
Radio ('wireless')		507
Tape or CD/DVD player; HiFi		508
Television		509
Sewing machine		511
Electric or gas stove; hot plate		513
Refrigerator		514
Bicycle		516
Motorcycle/scooter		517
Car		518
Upholstered chair, sofa set		522
Coffee table (for sitting room)		523
Cupboard, drawers, bureau		524
Lantern (paraffin)		525
Clock		527
Iron (for pressing clothes)		528

ENUMERATOR: RECORD PRIMARY RESPONDENT ID FOR MODULE B6: ____ ID

ENUMERATOR: RECORD END TIME FOR MODULE B6: ____ HOURS ____ MINUTES

MODULE B7: SHOCKS

ENUMERATOR: RECORD START TIME FOR MODULE B7: ____ HOURS ____ MINUTES

CODE	SHOCK	B701 During the last 12 months, was your household affected negatively by any of the following [SHOCK]? YES..1 NO...2 >> NEXT SHOCK	B702 Rank the three most significant shocks you experienced - **Most Severe (1), Second Most Severe (2), Third (3).**
101	Drought		
1101	Irregular Rains		
102	Floods		
1102	Landslides		
103	Earthquakes		
104	Unusually High Level of Crop Pests or Disease		
105	Unusually High Level of Livestock Disease		
106	Unusually Low Prices for Agricultural Output		
107	Unusually High Costs of Agricultural Inputs		
108	Unusually High Prices for Food		
109	End of Regular Assistance/Aid/ Remittances From Outside		
110	Reduction in the Earnings from Household		
111	Household (Non-Agricultural) Business		
112	Reduction in the Earnings of Currently		
113	Loss of Employment of Previously Salaried		
114	Serious Illness or Accident of Household Member(s)		
115	Birth in the Household		
116	Death of Income Earner(s)		
117	Death of Other Household Member(s)		
118	Break-Up of Household		
119	Theft of Money/Valuables/ Assets/Agricultural Output		
120	Conflict/Violence		
121	Other (Specify)		

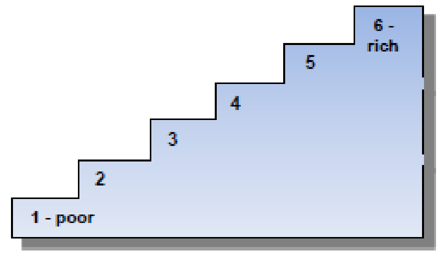
ENUMERATOR: RECORD PRIMARY RESPONDENT ID FOR MODULE B7: ____ ID

ENUMERATOR: RECORD END TIME FOR MODULE B7: ____ HOURS ____ MINUTES

MODULE B8: SUBJECTIVE WELFARE

ENUMERATOR: RECORD START TIME FOR MODULE B8: ____ HOURS ____ MINUTES

B801 Concerning your household's food consumption over the past one month, which of the following is true? It was less than adequate for household needs.... 1 It was just adequate for household needs.. ... . 2 It was more than adequate for household needs.... 3 *(NOTE THAT 'ADEQUATE' MEANS NO MORE OR NO LESS THAN WHAT THE RESPONDENT CONSIDERS TO BE THE MINIMUM CONSUMPTION NEEDS OF THE HOUSEHOLD.)*	B802 Imagine six steps, where on the bottom, the first step, stand the poorest people, and on the highest step, the sixth, stand the rich. SHOW THE PICTURE OF THE STEPS BELOW. On which step are you today?	B803 On which step are most of your neighbors today?	B804 On which step are most of your friends today?	B805 How many changes of clothes do you (HH HEAD) own? (NUMBER OF TROUSERS FOR MEN; SKIRTS/ DRESSES FOR WOMEN) NUMBER	B806 What do you (HH HEAD) sleep under in the cold season (July)? BLANKET & SHEETS. .1 BLANKET ONLY. . . . ..2 SHEETS ONLY . . . . .3 CHITENJE CLOTH. . . .4 FERTILIZER or GRAIN SACK . . . . .5 CLOTHES . . . . . . .6 NOTHING 7



ENUMERATOR: RECORD PRIMARY RESPONDENT ID FOR MODULE B8: ____ ID

ENUMERATOR: RECORD END TIME FOR MODULE B8: ____ HOURS ____ MINUTES

(THEN >> MODULE X)