Survey Ordering and the Measurement of Welfare[*]

Khandker Wahedur Rahman[†], Jeffrey Bloem[‡], and Marc F. Bellemare[§]

April 6, 2023

Abstract

Social and economic policy and research rely on the accurate measurement of welfare. In nearly all instances, measuring welfare requires collecting data via long household surveys that are cognitively taxing on respondents. This can lead to measurement error, both classical (i.e., noisier responses) and non-classical (i.e., biased responses). We embed a survey ordering experiment in a relatively short survey, lasting just over 75 minutes on average, by asking half of our respondents about their assets near the beginning of the survey (treatment) and asking the remainder of our respondents about their assets at the end of the survey (control). We find no evidence that survey ordering introduces classical or non-classical measurement error in either the number of reported assets or the reported asset value in the full sample. But in sub-samples of respondents who (i) are from larger (i.e., more than four individuals) households, or (ii) have low levels (i.e., fewer than six years) of education, we find evidence of differential reporting due to survey ordering. These results highlight important heterogeneity in response bias which, despite the null effect in the full sample, can be meaningful. For example, for respondents from larger households, placing the asset module near the beginning of the survey leads to a 23 percent increase in the total reported asset value relative to placing the same module at the end of the survey.

Keywords: Survey Design, Measurement Error, Poverty Measurement
JEL Classification Codes: C81, C83, I32, O12

[*] Authors are listed in reverse alphabetical order. The authors thank Raied Arman and Farhana Kabir for excellent research assistance in coordinating the fieldwork and data collection of the data used in this paper.

[†] University of Oxford, and BRAC Institute for Governance and Development, BRAC University. Email: khandkerwahedur.rahman@oxmfordmartin.ox.ac.uk.

[‡] International Food Policy Research Institute. E-mail: j.r.bloem@cgiar.org.

[§] Department of Applied Economics, University of Minnesota. E-mail: mbellema@umn.edu.

1 Introduction

"What gets measured gets improved," legendary management consultant Peter Drucker allegedly said. Even if the quote is apocryphal, it nevertheless conveys the idea that before one can tackle a problem, one must take stock of the extent of that problem, and this is just as true for management consultants as it is for the poor in low- and middle-income countries—the people in the world whose material lives are probably the most different from that of the typical management consultant's, and which could use the most material improvement.

As a result, social scientists whose ultimate goal is to reduce the extent of poverty have devised various ways of measuring poverty and keep improving on those ways. In places where the fiscal capacity of the state is limited, and where accurate income tax records are not available or are only available for the wealthier households in the population, the measurement of poverty relies on surveys that try to measure household income, expenditures, or assets as proxies for household welfare (Deaton, 1997). But given the level of detail involved in collecting precise values for those proxies for household welfare, household surveys can take a few hours of a survey respondent's time, and it is not unlikely that the longer a survey, the more respondents report mistaken information toward the end of that survey. For the researcher interested in using data from those surveys in applied research, this means that the data collected at the end of a survey might be more likely to suffer from measurement error, both classical (i.e., noisier responses) and non-classical (i.e., biased responses).

Is welfare measured with more error depending on the placement of the welfare measurement module within a survey? We answer this question by using a household's assets—both the number of assets as well as the total value of the household's assets—as a proxy for welfare.[1] Within a survey administered in Bangladesh, we randomly assign half of our survey

[1] While welfare is often measured by collecting data on household income or expenditures, see Sahn and Stifel (2003) on using assets as a measure of welfare in cases where the collection of income or expenditures data is costly. See Balboni, Bandiera, Burgess, Ghatak and Heil (2022) for recent work using assets as an outcome to study poverty traps.

respondents to answer questions about their assets early in the survey (treatment) and the remainder of our respondents to answer questions about their assets at the end of the survey (control). To assess whether the early asset module treatment introduces non-classical measurement error, we rely on a standard approach wherein we regress each measure of welfare on an early asset module treatment dummy. To assess whether the early asset module treatment introduces classical measurement error, we compare the treatment and control groups for each outcome of interest by conducting (i) Breusch-Pagan tests of group-wise heteroskedasticity after estimating the aforementioned regressions, and (ii) Kolmogorov-Smirnov tests of equality of distributions for the outcomes of those regressions.

We find no evidence that survey ordering introduces classical measurement error in the measurement of assets. Moreover, we find no evidence that survey ordering introduces non-classical measurement error in either the number of reported assets or the reported asset value in the full sample. Within sub-samples of respondents who (i) are from larger (i.e., more than four individuals) households or (iii) have low levels (i.e., fewer than six years) of education, we find evidence of differential reporting due to survey module placement. We also find suggestive evidence that respondents who are not the head of their household may be overreporting the value of their assets. These results highlight possibly important sources of heterogeneity in response bias due to survey module placement which, despite the null effect in the full sample, can be meaningful.

This paper is closely related to several recent papers that experimentally study the effects of survey fatigue and questionnaire design as well as response bias. First, using long (i.e., two- to three-hour) multi-module household surveys administered in Liberia and Malawi, Jeong, Aggarwal, Robinson, Kumar, Spearot and Park (2023) randomize the order of survey modules measuring assets and food consumption and find that an additional hour of survey time needed to reach a given question increases the probability that a respondent triggers a skip code by answering “No” to the question.[2] Second, and again using a long multi-module household survey administered in Ghana, Ambler, Herskowitz and Maredia (2021) randomize the order in which household members appear within the labor module of their survey, and find that moving a household member back by one position reduces their reported number of productive activities. Finally, Abay, Berhane, Hoddinott and Tafere (2022) study response bias in relatively short phone surveys administered in Ethiopia by using a study design similar to ours. The authors randomly assign a survey with the dietary diversity module early vs. late in the questionnaire and find that respondents receiving a late dietary diversity module report less dietary diversity.[3]

Our contribution is threefold. First, whereas the literature focuses on whether various survey modalities introduce non-classical measurement error (i.e., bias) in measurement, we test whether the within-survey placement of the welfare (here, assets) measurement module introduces both non-classical *and* classical measurement error. Given that proxies for welfare are often used as outcome variables in empirical work, knowing whether they suffer from classical measurement error matters for inference. Second, we document conditions under which the randomized placement of a survey module within a questionnaire leads to a null effect on average. Documenting these null effects is important both for the purpose of preventing publication bias (Stanley, 2005) and because our multi-module survey took respondents about 75 minutes to complete. This is considerably shorter than the multi-module survey administered by Jeong *et al.* (2023) and provides a useful proof-of-concept for the length of a multi-module household survey without measurable response bias from survey

[2] Using household surveys of similar length but administered in Kenya, Laajaj and Macours (2021) randomize the order of three modules (i.e., measuring cognitive, non-cognitive, and technical agronomic skills) and find no evidence of survey fatigue in their data. It should be noted, however, that the Laajaj and Macours (2021) study design is less focused on investigating bias due to response fatigue and more focused on testing the reliability of survey questions measuring different types of skills, which may be influenced by order effects and anchoring.

[3] In a related paper, Abate, De Brauw, Hirvonen and Wolle (2023) randomly assign households an in-person or phone survey and find that respondents to the phone survey reported 23 percent less consumption than respondents to the in-person survey.

fatigue. Third, we document heterogeneous effects in response bias due to survey module placement. Documenting these heterogeneous effects is important because they demonstrate that response bias due to survey fatigue can lead to non-classical measurement error even in a context where there is no measurable response bias in the full sample.

The remainder of this paper proceeds as follows. In the next section, we introduce our experimental design. In Section 3, we discuss both our main, pre-registered results and additional, exploratory results. We conclude in section 4 with recommendations for survey design and future research.

2 Experimental Design

We embed this survey experiment within the baseline survey of a larger experiment that aims to estimate demand for digital financial services in Bangladesh (Rahman and Bloem, 2020). We differentiate between the larger experiment (i.e., the study of demand for digital financial services) which, at the time of writing this paper, is ongoing, and the survey experiment (i.e., the study of survey ordering on the measurement of welfare) which is the focus of this paper.

In this survey experiment, we randomly assign respondents to two groups. Roughly half of the respondents (i.e., $n = 1,951$) whom we assign to the treatment group receive a survey where the module collecting information about household assets appears at the beginning of the survey; we refer to that group as the "early asset module" group in what follows. The remainder of the respondents (i.e., $n = 1,980$) whom we assign to the control group receive a survey where the module collecting information about household assets appears at the end of the survey. Table A.1, shown in the Supplemental Appendix, reports both basic summary statistics about our sample and balance tests that illustrate the validity of our randomization among observable variables. Our sample is 99 percent female, the average respondent is 38 years old, 93 percent of respondents are married, about half live in households with more

Table 1: Survey Module Order

Treatment	**Control**
Pre-screening	Pre-screening
Screening	Screening
Consent	Consent
Assets	Demographics
Demographics	Employment
Employment	Household Finances
Household Finances	Enterprise Outcomes
Enterprise Outcomes	Digital Financial Services
Digital Financial Services	Economic Empowerment I
Economic Empowerment I	Economic Empowerment II
Economic Empowerment II	Interpersonal Freedom
Interpersonal Freedom	Social Networks
Social Networks	Assets

Notes: Household survey module order for respondents in the treatment and control group. Statistics about the duration of active survey time are reported in Figure 1.

than four members, 34 percent of respondents are the head of their household, and about half have completed schooling up to about class six. These statistics are not significantly different between the treatment and control groups.

Table 1 lists the order of the survey modules for respondents in the treatment group and respondents in the control group. The key difference is whether respondents receive a survey with the asset module early in the questionnaire (i.e., treatment) or a survey with the asset module late in the questionnaire (i.e., control). Of course, the relative placement of other modules is also different between the treatment and control groups. The relative difference in the placement of those other, non-asset modules is much less than the relative placement of the assets module. Therefore, we assess the effect of assignment to the treatment group on reported assets. Each respondent, regardless of treatment status, received the same survey modules and questions.

Therefore, it is unsurprising to find that the distribution of the duration of active survey time is essentially identical between the treatment and control groups, as shown in Figure 1. On average, respondents took about 75 minutes to complete our survey, with the first percentile of the distribution completing the survey in about 40 minutes and the 99th percentile of the distribution completing the survey in about two hours and 15 minutes. Our survey is thus notably shorter than the 2.5 hour (on average) survey used by Jeong *et al.* (2023) in their survey experiment on survey fatigue. Figure A.1 in the Supplemental Appendix shows the order in which assets are listed in the asset module and the probability of reported asset ownership.

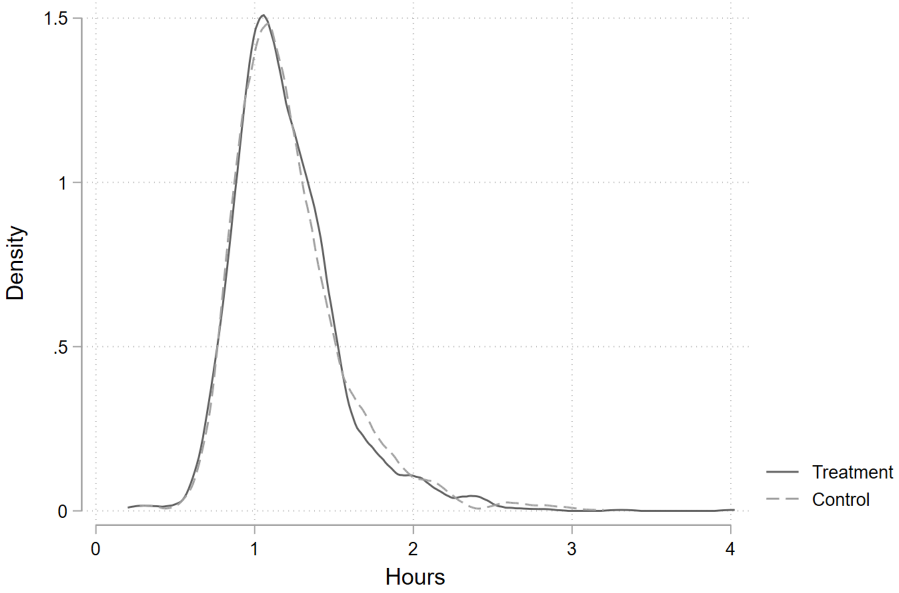

Figure 1: Kernel Density Estimate of the Duration of Active Survey Time

Notes: Epanechnikov kernel. Full sample mean = 1.21 hours, median = 1.14 hours, 1st percentile = 0.64 hours, and 99th percentile = 2.38 hours. Sample size = $3,931$. The average difference in duration by treatment status is not statistically significant. Regression results are shown in Table A.2.

Our analytical approach is straightforward. We compare reported assets between the treatment group and the control group to estimate the effect of receiving a survey with the asset module early in the questionnaire relative to receiving a survey with the asset module placed at the end of the questionnaire.

Our main estimation approach uses the following regression specification:

$$Y_{ij} = \alpha_j + \beta_j T_i + \gamma_i + \epsilon_{ij} \tag{1}$$

In equation (1), Y_i denotes the number of assets reported by the respondent from household i or the inverse hyperbolic sine of the reported value of each asset category j or the number of assets reported (Bellemare and Wichman, 2020), T_i is a dummy variable equal to one if household i is in the treatment group and equal to zero otherwise, and ϵ_{ij} is an error term with mean zero. As this survey experiment uses baseline data from a larger experiment (Rahman and Bloem, 2020), we also control for stratum fixed effects, γ_i, pertaining to treatment status in the larger experiment. The standard errors are clustered by survey experiment treatment status within centers (i.e., a micro-finance branch location), which represents the level of randomization from the larger experiment (Rahman and Bloem, 2020).

3 Results

This section presents our results for whether the placement of the asset module introduces both classical and non-classical measurement error. We first report our pre-registered results.[4] We then report the results of an exploratory analysis of treatment heterogeneity.

3.1 Pre-Registered Results

To test whether the placement of the asset module introduces non-classical measurement error, we start by estimating regressions aimed at assessing the differences in the number of assets reported and the inverse hyperbolic sine (i.e., asinh) of the reported value for each asset category. To test whether the placement of the asset module introduces classical

[4] Our pre-analysis plan is registered with the American Economic Association RCT registry and available here: https://www.socialscienceregistry.org/trials/10309.

measurement error, we then compare the mean of the squared residuals from those regressions with a Breusch-Pagan test of group-wise heteroskedasticity for each outcome. Additionally, we test whether the distribution of each outcome is equal across treatment and control groups using a Kolmogorov-Smirnov test to assess the robustness of our classical measurement error finding.

Table 2 reports results from estimating equation (1) with the number of assets reported as the dependent variable. The first column uses the raw number of assets as the dependent variable and the second column uses the natural log of the number of assets reported. In both columns, we find a null effect. In the first column, the coefficient indicates that respondents receiving a questionnaire with an early asset module report 0.15 more assets than respondents receiving a questionnaire with a late asset module, relative to a sample mean of about 11.5 assets reported. In the second column, we find that an early asset module leads to about 1.4 percent more assets reported. The estimates in both columns are relatively precise null effects, meaning they are small in magnitude and not statistically significant. Moreover, we conduct a Breusch-Pagan test of group-wise heteroskedasticity across the treatment and control groups. In both columns in Table 2, we fail to reject the null of no difference in the sum of squared residual by treatment status. Finally, a Kolmogorov-Smirnov test with a p-value of 0.784 shows that, along the entire distribution of the outcome variables used in Table 2, there is no difference between the treatment and control groups.

We now estimate differences in the reported value of each asset category. We observe 41 asset categories, but for many asset categories, many respondents report not owning any assets. We thus transform these values using the inverse hyperbolic sine transformation, which is log-like but allows retaining zero-valued observations. Figure 2 reports estimates of the effect of receiving a questionnaire with a relatively early asset module on the inverse hyperbolic sine of the reported value of assets for each category. We report these estimates from the lowest coefficient to the highest coefficient. Notably, none of the estimates are sta-

Table 2: Number of Assets Reported, Full Sample

	Number of Assets	Ln(Number of Assets)
Early Asset Module	0.155	0.014
	(0.154)	(0.014)
Observations	3,931	3,931
R-squared	0.002	0.002
Breusch-Pagan (p-value)	0.549	0.530
Stratum Fixed Effects?	Yes	Yes
Sample mean	11.58	2.41

Notes: Standard errors clustered by survey experiment treatment status within centers in parentheses *** p<0.01, ** p<0.05, * p<0.1

tistically significant at the 95 percent level. Moreover, this finding holds when we adjust for multiple hypothesis testing using the method developed by Benjamini, Krieger and Yekutieli (2006), as implemented by Anderson (2008).

We further conduct a series of Breusch-Pagan and Kolmogorov-Smirnov tests for each of the 41 asset categories and, after adjusting for multiple hypothesis testing, we find no difference in the squared residuals and no difference along the entire distribution of the outcome variables for each regression reported in Figure 2. By failing to reject the null of no classical or non-classical measurement error on the value reported for each of the 41 asset categories, these findings further support the null effect of survey ordering on the measurement of welfare, at least in our relatively short (i.e., 75 minute, on average) survey.

3.2 Exploratory Results

We now depart from our pre-registered results to examine some exploratory results that provide additional nuance and insights from our survey experiment. First, we report additional

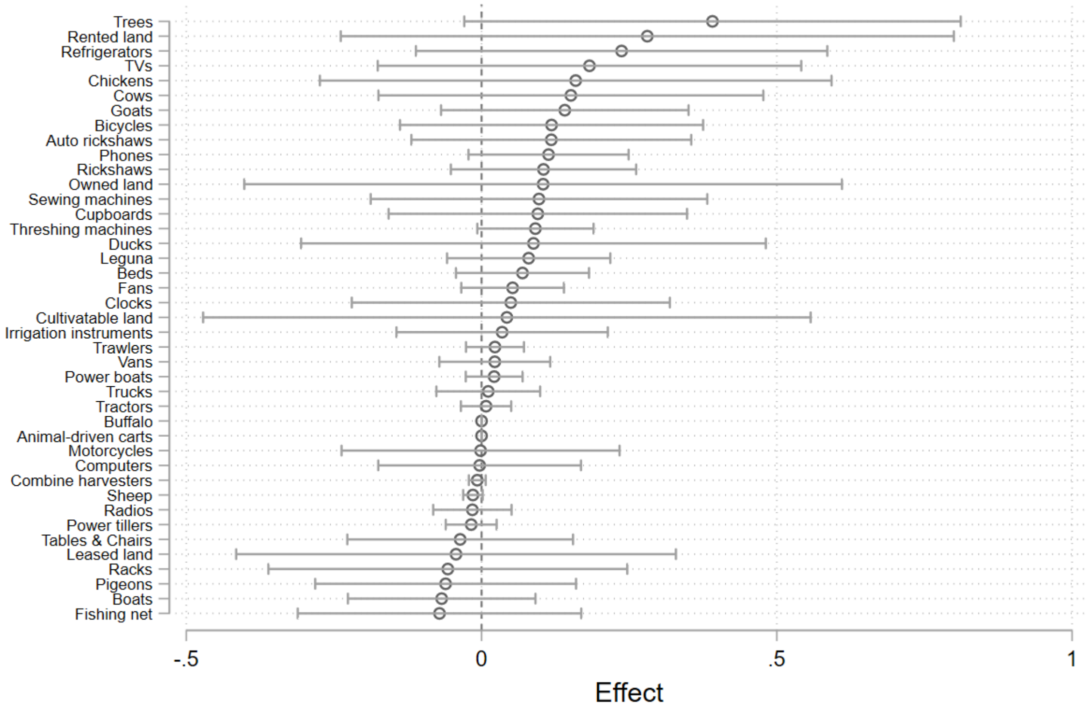

Figure 2: Effect on the Value Reported of Each Asset

Notes: This figure shows coefficient estimates with associated 95 percent confidence intervals. When we adjust for multiple hypothesis testing using the method developed by Benjamini *et al.* (2006), as implemented by Anderson (2008), none of these effects are statistically different from zero.

results from our full sample. Second, we document important dimensions of heterogeneity.

Additional Full Sample Results. Given the null effects reported in Table 2 on the total number of assets reported and in Figure 2 on the reported value of each asset, one may expect that we also find a null effect on the total reported asset value, and indeed this is what we find. Table A.2, shown in the Supplemental Appendix, reports the effect of receiving an early asset module on the natural log of total reported asset value. We find that our treatment led to a roughly 9-percent increase in the total reported asset value, but this estimate is imprecise, and thus not statistically significant. Nevertheless, this effect magnitude could represent a meaningful source of bias in some empirical settings.

Heterogeneity Analysis. Our second set of exploratory results focuses on effect heterogeneity within distinct sub-samples of our data. We explore three sub-samples: (i) respondents from households with more than four members, (ii) respondents who are not the head of their household, and (iii) respondents who have completed less than class six. As shown in Table A.1 in the Supplemental Appendix, these sub-samples each divide our sample roughly in half. Table 3 reports estimates of the effect of our treatment on the number of assets reported and the reported asset value for each of these sub-sample groups.

In panel A of Table 3, we report results for the sub-sample of respondents who live in households with more than four members, which effectively cuts our sample in half. It might be more challenging and cognitively taxing to accurately recall details about household assets when the household is relatively large. In the first column, we find that respondents receiving a questionnaire with an early asset module report 0.29 more assets than respondents receiving a questionnaire with a late asset module, relative to a sample mean of 12.11 assets reported. In the second column, we find that our treatment leads to about 2.8 percent more assets reported. In the third column, we find that receiving a questionnaire with an early asset module leads to a 23 percent higher total reported asset value. The estimate in this third column is statistically significant and represents notable response bias due to the placement of the asset module.

In panel B of Table 3, we report results for the sub-sample of respondents who are not the head of their household. Due to the nature of the "primary" randomized control trial that focuses on clients of local micro-finance branch centers, roughly 66 percent of our sample are not the head of their household. It might be that respondents who are not the head of their household have a more difficult time accurately recalling details about household assets because they exert less control over those assets. In the first column, we find that respondents receiving a questionnaire with an early asset module report 0.19 more assets than respondents receiving a questionnaire with a late asset module, relative to a sample mean of 11.61 assets reported. In the second column, we find that our treatment leads

Table 3: Number of Assets Reported and Total Asset Value, within Sub-Samples

	Number	ln(Number	ln(Total Asset
	of Assets	of Assets)	Value)
Panel A: Household Size (> 4 *members*)			
Early Asset Module	0.293	0.028*	0.233**
	(0.188)	(0.016)	(0.095)
Observations	2,000	2,000	2,000
R-squared	0.004	0.006	0.007
Sub-sample mean	12.11	2.45	14.24
Panel B: Household Head (= *no*)			
Early Asset Module	0.187	0.018	0.136
	(0.171)	(0.016)	(0.095)
Observations	2,608	2,608	2,608
R-squared	0.003	0.003	0.002
Sub-sample mean	11.61	2.41	14.00
Panel C: Low Education (< *class* 6)			
Early Asset Module	0.361*	0.034*	0.083
	(0.193)	(0.0182)	(0.116)
Observations	2,027	2,027	2,027
R-squared	0.006	0.005	0.003
Sub-sample mean	11.08	2.36	13.75
Stratum Fixed Effects?	Yes	Yes	Yes

Notes: Standard errors clustered by survey experiment treatment status within centers in parentheses *** p<0.01, ** p<0.05, * p<0.1

to about 1.7 percent more assets reported. In the third column, we find that receiving a questionnaire with an early asset module leads to a 14 percent higher total reported asset value. The estimates in each of these columns are not statistically significant, however, the estimate in the third column represents a meaningfully large response bias due to the placement of the asset module.

Finally, in panel C of Table 3, we report results for the sub-sample of respondents who have completed less than class six. It is possible that respondents who have completed relatively low levels of education might experience survey fatigue earlier than respondents who have completed higher levels of education. In the first column, we find that respondents receiving a questionnaire with an early asset module report 0.36 more assets than respondents receiving a questionnaire with a late asset module, relative to a sample mean of 11.08 assets reported. In the second column, we find that our treatment leads to about 3.4 percent more assets reported. In the third column, we find that receiving a questionnaire with an early asset module leads to an 8 percent higher total reported asset value. The estimates in the first two columns are statistically significant at the 0.1 percent level, but the estimate in column three is not statistically significant.

4 Conclusion

Does the within-survey placement of a module aimed at measuring household welfare, as proxied here household assets (Balboni *et al.*, 2022; Sahn and Stifel, 2003), introduce measurement error, either classical (i.e., noisier responses) or non-classical (i.e., biased responses)? We answer this question by randomizing the placement of the asset module earlier in the questionnaire (treatment) or at the end of the questionnaire (control) within a relatively short (i.e., 75-minute, on average) household survey in Bangladesh.

We find no evidence that survey ordering introduces classical measurement error in the measurement of assets. Additionally, we find no evidence that survey ordering introduces non-classical measurement error in the measurement of assets in the full sample. Here it is important to again emphasize that our survey was short relative to other surveys included in the survey fatigue literature, and therefore these null results provide a useful proof-of-concept for the length of a multi-module household survey without measurable response bias, at least on average.

We do, however, find heterogeneous results in various sub-samples. Respondents who are from larger (i.e., more than four individuals) households and respondents who have low levels (i.e., less than class six) of education are likely to report more assets or a greater value of total assets when the asset module is placed early in the survey questionnaire.

One limitation of our approach is that while our research design allows testing whether there are systematic differences in the mean and variance of our proxy for welfare, it does not allow testing which of the two versions of the questionnaire—the early asset module version or the version with the asset module at the end of the questionnaire—is closer to the truth. Although there are good reasons to believe that the dominant mechanism at play is survey fatigue in our exploratory analysis, where we find systematic differences in assets between our treatment and control groups in specific sub-samples, we cannot rule out other mechanisms. Future research should aim to test which of early or late asset modules are more likely to generate answers closer to the truth, and thus to test for the precise mechanism whereby reported assets (or other proxies for welfare) differ. One way to do this could be to randomize both the within-survey placement of survey sections and the length of the survey (say, by introducing additional middle sections), and then test whether survey length is a mediator in the relationship between survey section placement and reported assets.

Finally, our results highlight a potential trade-off when it comes to designing surveys. If the placement of a given module influences the quality of the data measured by this module, then there may also be non-classical measurement error embedded in survey modules coming relatively later in the questionnaire. Ostensibly, this implies that researchers should design surveys with the most important modules at the beginning of the questionnaire. In practice, however, most information included in any given survey is likely to be important for some purpose or another. Thus, ordering survey modules based on their relative level of importance may not be a reasonable task. In that case, a potential solution is to divide the survey into multiple sessions—or administer the survey in short time periods across multiple days.

References

Abate, G. T., De Brauw, A., Hirvonen, K. and Wolle, A. (2023) Measuring consumption over the phone: Evidence from a survey experiment in urban ethiopia, *Journal of Development Economics*, **161**, 103026.

Abay, K. A., Berhane, G., Hoddinott, J. and Tafere, K. (2022) Respondent fatigue reduces dietary diversity scores reported from mobile phone surveys in ethiopia during the covid-19 pandemic, *The Journal of Nutrition*, **152**, 2269–2276.

Ambler, K., Herskowitz, S. and Maredia, M. K. (2021) Are we done yet? response fatigue and rural livelihoods, *Journal of Development Economics*, **153**, 102736.

Anderson, M. L. (2008) Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects, *Journal of the American statistical Association*, **103**, 1481–1495.

Balboni, C., Bandiera, O., Burgess, R., Ghatak, M. and Heil, A. (2022) Why do people stay poor?, *The Quarterly Journal of Economics*, **137**, 785–844.

Bellemare, M. F. and Wichman, C. J. (2020) Elasticities and the inverse hyperbolic sine transformation, *Oxford Bulletin of Economics and Statistics*, **82**, 50–61.

Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006) Adaptive linear step-up procedures that control the false discovery rate, *Biometrika*, **93**, 491–507.

Deaton, A. (1997) *The analysis of household surveys: a microeconometric approach to development policy*, World Bank Publications.

Jeong, D., Aggarwal, S., Robinson, J., Kumar, N., Spearot, A. and Park, D. S. (2023) Exhaustive or exhausting? evidence on respondent fatigue in long surveys, *Journal of Development Economics*, **161**, 102992.

Laajaj, R. and Macours, K. (2021) Measuring skills in developing countries, *Journal of Human resources*, **56**, 1254–1295.

Rahman, K. W. and Bloem, J. R. (2020) Digital finance and economic empowerment: Experimental evidence on the role of transaction costs, *BRAC BIGD Project*.

Sahn, D. E. and Stifel, D. (2003) Exploring alternative measures of welfare in the absence of expenditure data, *Review of income and wealth*, **49**, 463–489.

Stanley, T. D. (2005) Beyond publication bias, *Journal of economic surveys*, **19**, 309–345.

Supplemental Appendix

The Supplemental Appendix includes the following additional results.

- Tables
  - Table A.1 reports basic summary statistics and balance tests between the treatment and control groups.
  - Table A.2 reports regression results on the effect of our treatment on the duration of active survey time and the natural log of total asset value for the full sample.
- Figures
  - Figure A.1 reports the probability of asset ownership for each asset category in the order in which the categories appear within the assets module.
  - Figure A.2 shows the estimated effects on the value reported of each asset category for the sub-sample of respondents who live in households with more than four members.
  - Figure A.3 shows the estimated effects on the value reported of each asset category for the sub-sample of respondents who are not the head of their household.
  - Figure A.4 shows the estimated effects on the value reported of each asset category for the sub-sample of respondents who have an education of less than class six.

Appendix Tables

Table A.1: Summary Statistics

	Treatment	**Control**	**Difference** (C-T)
Age	37.8	38.1	-0.303
Married	0.93	0.93	0.002
Large household ($> 4\ members$)	0.51	0.50	0.008
Household head	0.34	0.34	-0.003
Low education ($< class$ 6)	0.51	0.53	-0.019
Larger Treatment 1	0.34	0.33	0.005
Larger Treatment 2	0.17	0.17	0.001
Larger Treatment 3	0.16	0.17	-0.008
Larger Treatment 4	0.16	0.37	-0.009
Larger Treatment 5	0.17	0.16	0.010
N	1,951	1,980	

Notes: This table presents summary statistics and balance tests. All differences between the treatment and the control group are small and not statistically significant. Our sample also is almost entirely (i.e., 99 percent) female.

Table A.2: Other Outcomes of Interest

	Survey Duration (Hours)	ln(Total Asset Value)
Early Asset Module	-0.013	0.094
	(0.011)	(0.087)
Observations	3,931	3,931
R-squared	0.038	0.001
Stratum Fixed Effects?	Yes	Yes

Standard errors clustered at the center level in parenthesis
*** p<0.01, ** p<0.05, * p<0.1

Appendix Figures

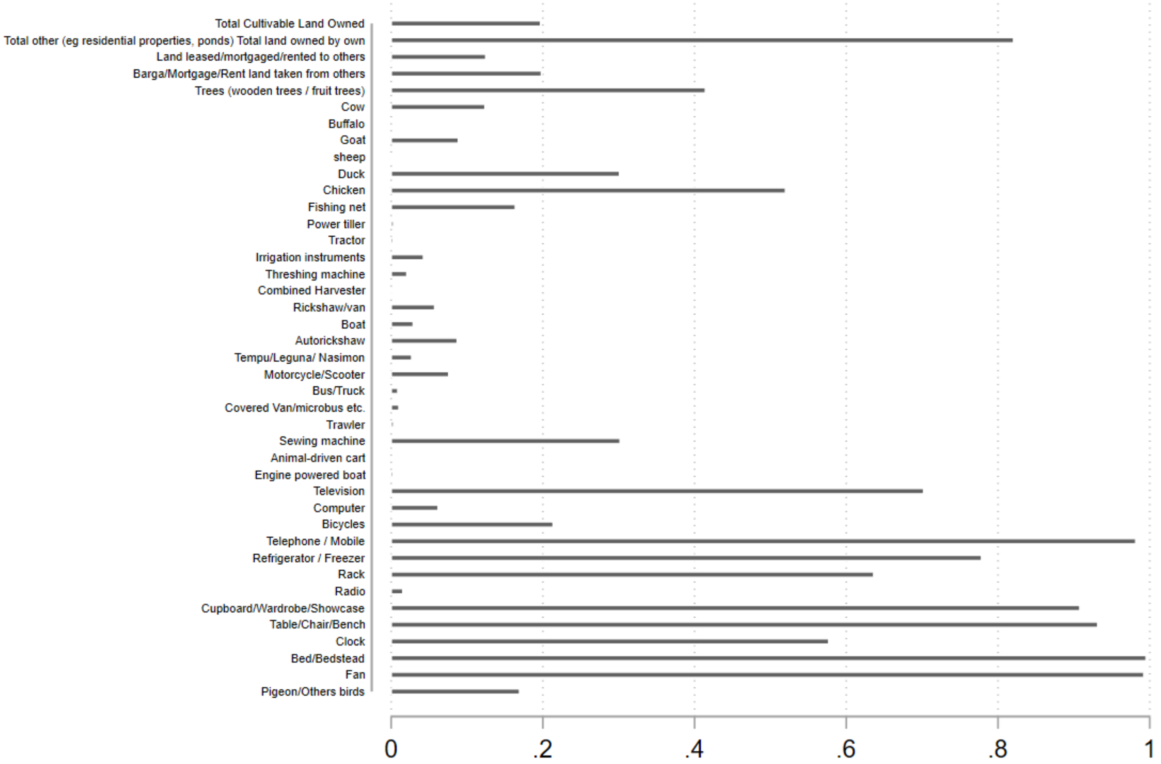

Figure A.1: Probability of Asset Ownership

Notes: This figure shows the probability of asset ownership for each asset category in the order in which these categories appear within the assets module.

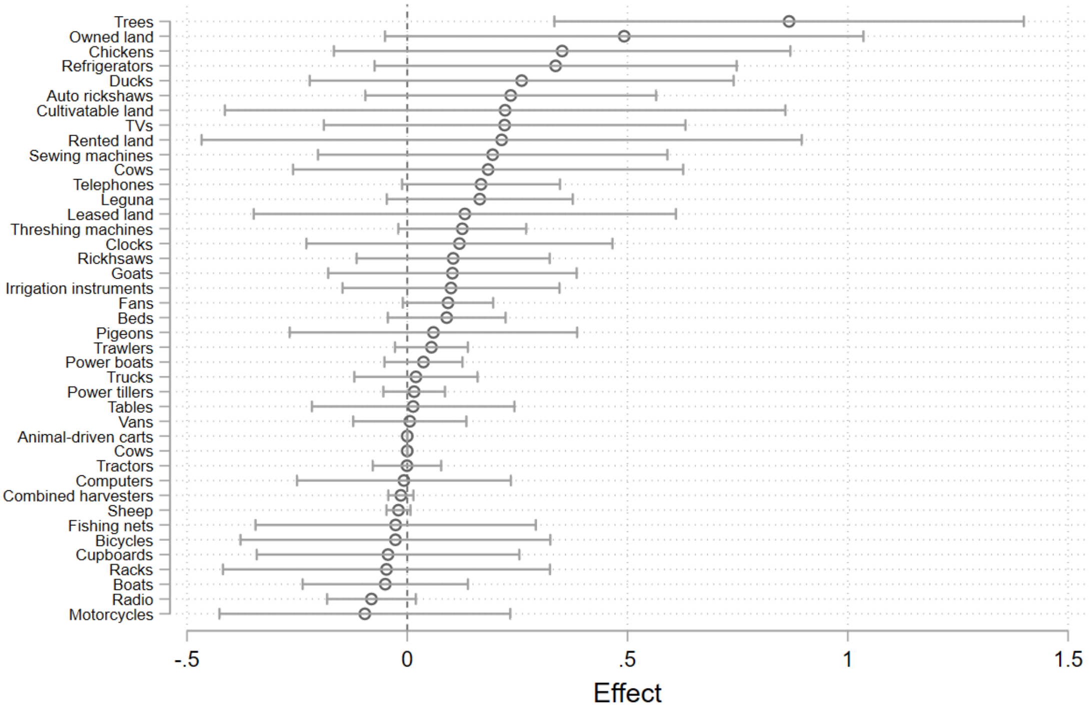

Figure A.2: Effect on the Value Reported of Each Asset—Household Size (> 4 *members*)

Notes: This figure shows coefficient estimates with associated 95 percent confidence intervals. When we adjust for multiple hypothesis testing using the method developed by Benjamini *et al.* (2006), as implemented by Anderson (2008), none of these effects are statistically different from zero.

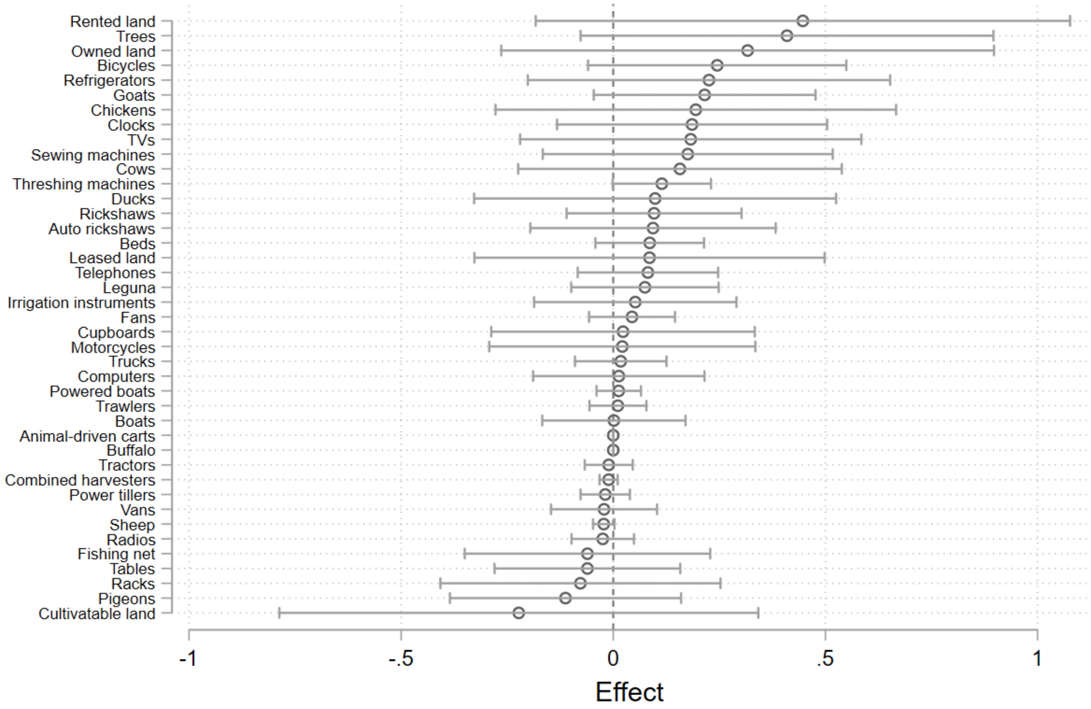

Figure A.3: Effect on the Value Reported of Each Asset—Household Head (= *no*)

Notes: This figure shows coefficient estimates with associated 95 percent confidence intervals. When we adjust for multiple hypothesis testing using the method developed by Benjamini *et al.* (2006), as implemented by Anderson (2008), none of these effects are statistically different from zero.

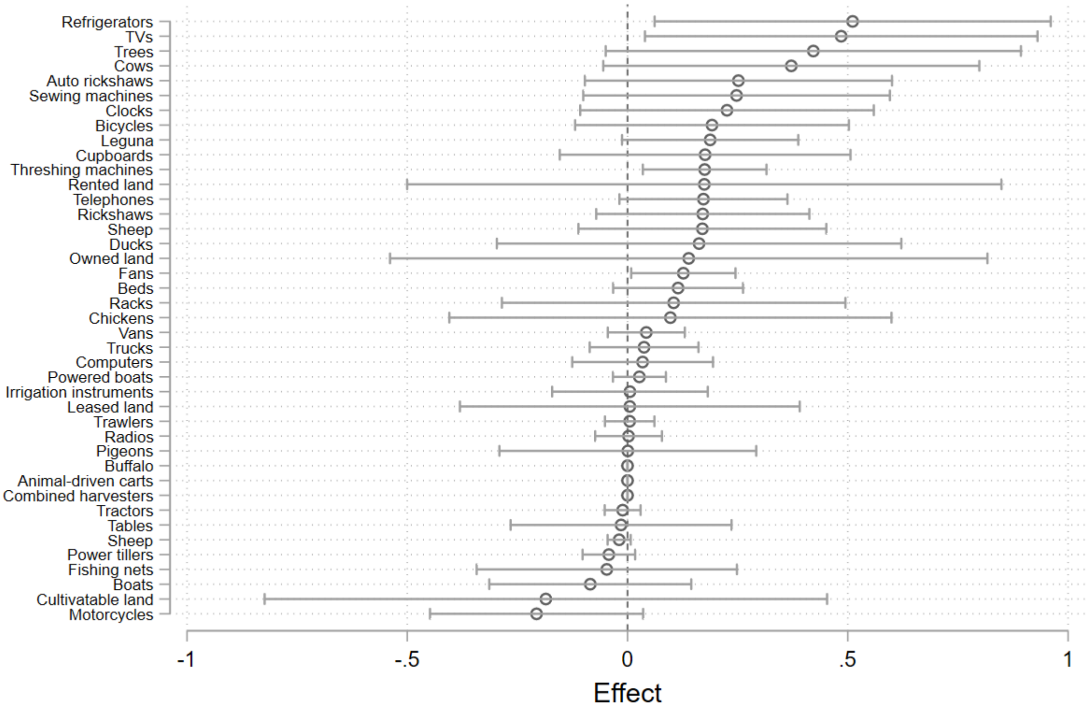

Figure A.4: Effect on the Value Reported of Each Asset—Low Education (< *class* 6)

Notes: This figure shows coefficient estimates with associated 95 percent confidence intervals. When we adjust for multiple hypothesis testing using the method developed by Benjamini *et al.* (2006), as implemented by Anderson (2008), none of these effects are statistically different from zero.