Are We Done Yet?
Response Fatigue and Rural Livelihoods

Kate Ambler[†], Sylvan Herskowitz [*†], and Mywish Maredia[‡]

[†]International Food Policy Research Institute
[‡]Michigan State University

February 1, 2021
Link to most recent version

Abstract

Effective policy requires an accurate understanding of peoples' livelihoods activities. The data for this evidence is often generated via lengthy surveys where designated respondents provide information about their household members. This burden on respondents may lead to both losses and biases as they grow fatigued during the interview. We test these hypotheses with an experiment in rural Ghana where we randomize individual household members' position in the labor module. We find that moving a household member back by one position reduces their reported number of productive activities by 2.2% with average aggregate losses of 8%, or approximately one out of every twelve activities. For women and youth, losses are closer to one in nine. These biases result from both differential exposure to response fatigue (being positioned later in rosters) and differential vulnerability (greater impacts conditional on position). These results have important implications for data quality across many settings and topics.

Keywords: Labor, Response Fatigue, Gender, Youth, Survey Methods, Proxy Reporting
JEL Codes: O1, J2, C8, Q1

[*]Send correspondence to: s.herskowitz@cgiar.org. This project was funded by 3ie and the CGIAR Research Program on Policies, Institutions, and Markets, and was approved by the IFPRI Institutional Review Board. We thank Mike Murphy for his project and data management as well as the team at Innovations for Poverty Action Ghana for project implementation, research assistance, and data collection, particularly Usamatu Salifu, Nicole Gargano, Salifu Amadu, Hassan Moomin, Federica Di Battista, and Madeleen Husselman. We also appreciate valuable feedback provided by Kathleen Beegle, Fiona Burlig, Susan Godlonton, Erin Kelley, Greg Lane, and seminar participants at IFPRI.

1 Introduction

Well-designed poverty alleviation programs and policies require an accurate understanding of peoples' livelihoods strategies and labor activities. In developing country settings, this understanding often relies on lengthy, multi-topic household surveys that document labor supply by collecting information on the productive activities of adult household members (Grosh et al., 2000). A recent increase in policy focus on work opportunities for women and youth has driven a stream of studies characterizing the different labor activities of these groups.[1] Another related body of research using this data explores the diversification of rural livelihood strategies, finding that off-farm activities are increasingly influential in the reduction of rural poverty.[2] The usefulness of these important areas of work relies fundamentally on the quality of the underlying data.

In this paper, we examine an important and understudied factor affecting the quality of existing micro data on labor supply: response fatigue. We conduct an experiment in which we randomize the order of household members in the labor module. The respondent thus reports on the labor activities of other members in random order, instead of the order in which those members were listed in the household roster. We find that those who are listed later, and thus exposed to greater response fatigue, are reported to have participated in fewer work activities than those positioned earlier. Our analysis also shows that this fatigue-induced underestimation can lead to substantial bias in the resulting data. This bias enters in two ways. First, we demonstrate differential *vulnerability* to fatigue, whereby losses for women and youth are greater than those for men and older people, conditional on roster position. And second, we observe differential *exposure* to fatigue that results from women and youth being systematically listed later in household rosters.

This paper uses data from a survey in rural Ghana. Following a standard survey design, households' primary respondents were asked to report details about their work

[1] For recent examples see Krumbiegel et al. (2020), Betcherman and Khan (2018), Van den Broeck and Kilic (2019), Bridges et al. (2011), and Klasen and Lamanna (2009).

[2] Recent examples include Dzanku (2020), Asfaw et al. (2019), Yeboah and Jayne (2018), Imai et al. (2015), Himanshu et al. (2013), Djurfeldt (2013), Haggblade et al. (2010), Ellis (1998), and Ellis and Freeman (2004).

activities as well as those of each household member age fourteen and above.[3] Considering one member at a time, respondents were asked to list the primary activity over the last year, the second main activity over the last year, and the primary activity over the last week for each member. Each affirmative answer triggered a set of follow-up questions to provide further details on the reported activity. Through this iterative process, it is quickly evident to respondents that by reporting fewer work activities they can decrease their response burden, reducing the number of questions they are asked and the time needed to complete the module. If respondents became fatigued from the repetitive questions, we would expect them to report fewer labor activities as they progressed through the list of eligible household members.

By randomizing the order in which household members are considered, instead of following the initial household roster (as conventionally done), we can estimate the causal effect of being exposed to greater response fatigue. We find a 2.2% reduction in the number of tasks reported for each spot later that an individual is positioned in the response order, controlling for household fixed effects and individual-level characteristics.

These effects of response fatigue vary across groups, suggesting differential vulnerability to response fatigue. Moving back in household rosters has a greater impact for female and younger household members, 3% of the mean for both groups per position. While we randomized the order of our labor module, most surveys follow the order of the initial household roster. A second mechanism by which response fatigue could bias the resulting data is if different groups are systematically positioned later in household rosters and thus exposed to greater fatigue. We observe this directly in our household orderings (and document similar patterns in other data sources as well). In our data, women are positioned slightly later than men in the household roster, 0.2 positions on average, whereas youth age 14-24 are located 1.4 positions later than adults age 35-59.

We combine these two insights and use these estimates with the original, non-random, household listings to generate predictions of the aggregate effects of response fatigue for different sub-populations. We do this by calculating counterfactual "no fatigue" estimates of labor activities for everyone in the sample as well as counterfactual estimates

[3] This is similar to the approach used in Living Standards Measurement Study surveys (Schaffner, 2000)

that incorporate the effect of fatigue as it would have impacted responses had the original, non-random ordering been followed. We do this first with a simple model that holds the impact of fatigue equal for all types of household members (but is still influenced by household roster position) and then for two other versions of the model, one that allows for heterogeneity by gender, and a second allowing for heterogeneity by age group. We estimate that overall losses in terms of unreported labor activities are just under 8%. Allowing for heterogeneity by gender, we find losses for women of over 11%, more than 2.5 times larger than the losses for men. Finally, we find that youth (age 14-24) and younger adults (age 25-34) both have losses of approximately 11%, more than five times greater than the losses of older adults (age 35-59). These results illustrate the importance of both differential exposure and vulnerability to response fatigue in generating biases in the resulting data.

While the reported estimates are based on our rural sample in Ghana, our findings raise concerns that extend both outside this setting and beyond labor modules. Respondents in any location are likely to become tired or bored by similarly structured labor modules. This form of fatigue is also likely to impact responses on entirely different topics, such as household consumption, expenditures, agricultural input use, or birth histories, all of which frequently use modules that follow similar iterative structures where affirmative responses for a given item trigger a set of follow-up questions. All of these modules are further likely to be impacted by non-random ordering in the sequence of items or people and therefore create systematic differences in exposure to response fatigue for different items. Certain types of modules may be even more vulnerable, such as consumption which frequently asks about more than one hundred different food items, each with multiple follow-up questions. With cereals often listed first and sweets towards the end we may be systematically undercounting consumption of unhealthy foods with major public health implications.[4]

Research on survey methodology has grown rapidly in recent years, propelled by the

[4]Beegle et al. (2012b) address the length of consumption modules by randomizing respondents into a long module, a short module with collapsed categories, or a short module focused on representative categories. They find that a short module with representative categories performs similarly to a long module, but they do not address the issue of data quality decay with the module.

proliferation of household surveys in developing countries.[5] Topics addressed in this work have included the level of detail in questions, selection of the household's primary respondent, the use of screening questions, different recall windows, household definitions, and the use of high frequency checks.[6] Galdo et al. (2020), Desiere and Costa (2019), and Comblon et al. (2015) provide examples of how other design choices such as question phrasing and recall windows can induce biases by age and gender in survey responses, similar to our findings. In other work on the use of proxies in labor modules, assessments have been mixed. Dillon et al. (2012) do not find evidence that proxy reporting impacts reported measurement of labor participation among children. However, Bardasi et al. (2011) show that use of proxies may reduce reported levels of male employment, though not females. And Desiere and Costa (2019) recently found that proxy reporting leads to under reporting of labor activities in a sample from Nigeria, but the mechanisms are left unexplored. Our work adds to this expanding literature.

Despite this growing research on survey design, the issue of response fatigue has received less attention. In a review of lessons from 15 years of experience with the World Bank's Living Standards Measurement Study (LSMS) (Grosh et al., 2000), discussions on household roster construction (Glewwe, 2000) and labor modules (Schaffner, 2000) acknowledge concern that fatigue could lower data quality over the course of interviews, but do not examine it directly nor discuss the implications for creating biases in the resulting data. In one exception, Laajaj and Macours (forthcoming) randomize the order of survey modules to test the effect of fatigue on a set of skills assessments in a study in Kenya but do not find meaningful effects. In other fields, such as in health and criminology, concern has been raised about fatigue, with work showing that data quality can deteriorate over the course of an interview and that attrition from future survey rounds is

[5] See for example the special issue on measurement in the Journal of Development Economics (McKenzie and Rosenzweig, 2012).

[6] Examples of research on question type and detail include Bardasi et al. (2011), Deininger et al. (2012), Langsten and Salen (2008), Comblon et al. (2015), and Benes and Walsh (2018). For examples of research on screening questions see Martin and Polivka (1995), Serneels et al. (2016), and Fox and Pimhidzai (2013). For examples related to recall windows see Beegle et al. (2012a), Heath et al. (2020), Das et al. (2012), Deininger et al. (2012), Gaddis et al. (2020), and Arthi et al. (2016). See Beaman and Dillon (2012) for work on household definitions. For work on high frequency checks see Caeyers et al. (2012) and Fafchamps et al. (2012).

more likely when baseline surveys are longer.[7] Our paper contributes to this literature by linking response fatigue to the understanding of rural livelihoods and showing statistically significant and economically meaningful losses in recorded labor activities. We also add to the literature by estimating losses for different sub-groups along with evidence on how two distinct mechanisms, differential exposure and vulnerability, can lead to substantial biases in the resulting data. Our results speak not only to the methodological literature on survey design, but also have implications for any research using similarly structured data collection.[8]

This paper proceeds by first providing background on the setting, data, and empirical strategy in Section 2. In Section 3 we present our main results. In Section 4 consider a number of important extensions and their implications while Section 5 concludes with a brief discussion.

2 Data and Empirical Strategy

We use data from a household survey conducted in Northern Ghana between April and June of 2019. The sample covered 12 districts in four regions. Respondents were members of farmer business organizations organized by the Ghana Agricultural Sector Investment Programme (GASIP) for the purpose of involvement in their agricultural programs. Sixty-six of these organizations, one per village, were included in this survey, serving as the baseline for a field experiment studying the adoption of conservation agriculture techniques. The survey included a detailed labor module in addition to others on agricultural activities, production practices, and knowledge.[9] Given the purpose of the original study, the survey targeted program participants as each household's primary respondent. In total, 1,106 households were interviewed as part of the study.

The survey's labor module was modeled after the Ghana Living Standards Survey and the Uganda National Panel Survey. It followed a structure common to LSMS and other

[7]For examples of data quality deterioration, see Bradley and Daly (1994), Holbrook et al. (2007), Hess et al. (2012), Galesic and Bosnjak (2009), Roberts et al. (2010), Egleston et al. (2011), and Sharp and Frankel (1983). For examples of attrition see Rolstad et al. (2011) and Hart et al. (2005)

[8]Notable examples include Dolislager et al. (2020), Yeboah and Jayne (2018), and Davis et al. (2010).

[9]See Ambler et al. (2020) for further details and results of the study.

multi-topic surveys implemented by national statistics offices in developing countries. In the labor module, the respondent is asked to report the labor activities of each household member age 14 and above. Progressing one member at a time, the respondent is asked if that member participated in any productive activities over the last 12 months. If so, they are asked a series of questions to identify and describe their primary activity over that period, gathering information on type of work, time use, and earnings. They are then asked if they had any secondary activities for that period and, if so, the follow-up questions are repeated. Finally, they are asked to identify their primary productive activity over the last seven days, whether this was one of the activities already described, and if not, again asked to provide further details. The respondent then repeats this sequence for the next household member. Each individual is therefore characterized as having participated in between 0 and 3 unique activities.

This design is intended to capture details of the most important productive activities for each individual over the past year, while also capturing primary activities over a seven day period to mitigate recall bias. However, respondents quickly learn from this structure that each new work activity they acknowledge leads to a full set of follow-up questions. The median completion time for the labor module was 18.5 minutes, compared to a median completion time of 104 minutes for the full interview.[10] The median time spent per eligible household member was five minutes.[11]

Our survey uses proxy response for most observations so that the primary respondent reported about the activities of the members in their household. However, instructions in the survey allowed own response (or for the respondent to confer) if that member was readily available. In general, household surveys vary in the extent to which they allow for or prioritize proxy versus self reports, but our survey is common in employing an approach that allows for both. 17% of individuals in our analytical sample are recorded as having reported for themselves (or been conferred with), an incidence that is low, but

[10] The labor module was the third section of the survey, and the median time elapsed prior to beginning the labor questions was 20.3 minutes.

[11] Conditional on the number of eligible household members, an additional member for which positive productive activities are reported increases total module length by approximately four minutes. The survey did not record the time spent on each individual, only the time spent on the entire module.

within the range of other similar surveys.[12] Implications of this allowance for our results are discussed further in section 4.2.

We hypothesize that response fatigue increases as respondents progress through the survey and with each repetition of the labor module. As a result, respondents may report fewer distinct work activities for family members listed later in household rosters in order to avoid additional associated follow-up questions. Labor modules typically ask about household members in the order in which they were constructed at the time of the initial roster listing. Similar to the guidance provided in most household surveys, the respondent in this survey was asked to list themselves first, followed by their spouse (if they had one), other adults in the house, and then children.[13] While these instructions are offered to ensure that household members are not unintentionally omitted, they also create patterns across households in which roster order is systematically related to gender and age. However, strict adherence to this ordering is not enforced (or observed in the data) and even with this guidance, respondents retain considerable discretion while filling out the roster. As such, household roster orders may also capture biases or heuristics of the respondent such as listing higher earners or people with greater stature before others. Given the range of factors that likely influence the order of household listings in this survey and many others, we cannot typically distinguish if correlations between work activities and listing order reflect real differences in household labor contributions or if they are driven or distorted by response fatigue.

In order to estimate the impact of response fatigue on reported labor activities, we randomized the order of the household roster (excluding the respondent) deployed in the labor module. Respondents reported information on their own labor activities first and then the module was repeated for each eligible member. Because the respondent's own position was not randomized, they are excluded from the analysis. Although they were included in the module, we additionally exclude those sixty years old and above (9% of the sample) from our analysis. This is done to focus on those who are below the

[12]The review by Desiere and Costa (2019) of the labor module data of LSMS surveys across six countries in Africa indicates that the rate of self-report responses ranged from 15% in Mali to 76% in Nigeria.

[13]Respondents were also told that a member of the household is someone who had slept in the respondent's house for 30 days consecutively or 60 days non-consecutively within the last 12 months, and shares food and other resources from a common source.

retirement age and who are still capable of working.[14] Finally, our preferred specification uses household fixed effects and therefore requires at least two non-respondent household members. Households with fewer than two members aged 14 to 59 (excluding the respondent) are therefore dropped from the analysis.[15]

2.1 Summary Statistics and "Balance"

Table 1 presents characteristics of the households and individuals included in the analysis sample. Households vary greatly in size but are large on average, with over eleven total members and six who were eligible for the labor module (14 or older) including the respondent. The large size of these households is both a result of the typical household structure in Northern Ghana as well as the selection criteria on our sample discussed in the previous section.[16] Approximately 30% were polygamous. 51% of respondents are female, 33% have ever attended school, and their average age is 42-years-old.

Panel B shows the non-respondent household members included in the analysis. 53% are female, they are 26-years-old on average, and 56% are literate with 36% currently in school. The majority of household members are children (45%), spouses (29%), siblings (12%), or parents (4%) of the respondent.[17] Household members are reported to have an average of 0.74 distinct work activities. 44% have no reported activities, 39% one activity, 18% two activities and just 0.3% had three distinct job activities listed. Among those who are working, 84% participate in household farm work as one of their listed activities. 15% participate in a household business and 25% engage in some form of hired, wage work. The mean position in the labor module for non-respondent household members is 3.8.

[14] Ghana's official retirement age is 60 years old and labor force participation drops sharply at this age. Additionally 44% of seniors in our sample are reported to be physically incapable of working, whereas just 1.6% of household members below 60 are similarly incapacitated. We show robustness of our main results to including seniors in the analysis in Appendix Table A.2.

[15] Appendix Table A.1 compares samples resulting from these two selection criteria.

[16] While these are large households, even by Ghanaian standards, they are only slightly larger, with one additional member on average, than rural Ghanaian households in the 2017 Ghana Living Standards Survey sample for the same districts and same selection criteria.

[17] Summary statistics of the full sample (before the sample criteria were applied) are shown in Appendix Table A.1. Characteristics are broadly similar but the analysis sample has larger households and is younger on average because the eligibility criteria drops people 60 and over and households with fewer than two people aged 14 to 59 other than the respondent.

If response fatigue accrues over the course of an interview, one's position in the household roster could impact exposure to fatigue. Column 1 of Table 2 shows that roster positions in the initial household listings (prior to randomization) are strongly correlated with personal characteristics. The order variable is the household member's position among other household members in the labor module, ranging from 1-14, where 1 represents the first member listed excluding the primary respondent.[18] We present the partial correlations of each individual's gender, age group, student status, and relationship to the primary respondent with their order in the household roster, controlling for household fixed effects. Women are listed later by an average of 0.2 positions ($p<.05$). Younger individuals are also listed later in the roster with individuals in the top age group, age 35-59, listed an average of 1.4 positions earlier than those aged 14-24 ($p<.01$), and those aged 25-34 listed 0.4 positions earlier than the youngest group ($p<.01$). We also observe strong patterns with relationship to the respondent (spouse of the respondent is the omitted category). Finally, we note that those who report their own labor are listed 0.9 positions earlier than those who are not, and that student status is not significantly related to roster position.

In column (2) of Table 2 we repeat the same exercise, but use the randomized order assignment as the dependent variable in place of the original listing order. Because of the randomization, we do not expect significant patterns and this analysis therefore serves as a balance check. All coefficients are small and only one (out of nine) is statistically significant at the 10% level.

2.2 Empirical Approach

We expect response fatigue to manifest in respondents reporting fewer productive activities as the module is repeated for each household member. Our primary outcome of interest is the number of distinct productive activities listed during the module for that household member. Using the randomized ordering of household members in the labor module, we estimate the causal impact of response order on reported labor activities

[18] The top 1% of both the listed and randomized household orders are winsorized at position 14 to reduce the influence of a long right tail in the distribution of household sizes.

using the following regression specification:

$$jobs_{i,h} = \beta_0 + \beta_1 order_i + \beta_2 fem_i + \beta_3 inschool_i + \beta_4 self_i + \beta_5 Age14\_24_i + \beta_6 Age25\_34_i + \gamma_h + \psi_r + \epsilon_{i,h} \tag{1}$$

$jobs_{i,h}$ is the number of unique work activities listed for individual, i, from household, h, ranging from 0-3. $order_i$ is the individual's randomly assigned order number (1-14) for this individual.[19] fem_i indicates the gender of the individual. $inschool_i$ controls for whether the individual is currently in school. $self_i$ indicates if this household member was conferred with or reported for themselves during the labor module. $Age14\_24$ indicates that the individual was aged 14 to 24. $Age25\_34$ indicates the 25-34 age group, with age 35-59 as the omitted category. ψ_r are a set of relation to respondent fixed effects. γ_h are a set of household fixed effects. These household fixed effects control for fixed characteristics of both the primary respondent and household and also constitute the strata of the randomization given that order position was randomly assigned within household. $\epsilon_{i,h}$ is the error term, clustered at the household level.

3 Results

3.1 Reduced Form Effects and Heterogeneity

Table 3 presents estimates of the impact of response fatigue on total jobs reported. The first three columns build up to our preferred specification from equation (1). Column (1) shows results from a regression of the total number of jobs on response order, controlling only for fixed effects for the number of eligible household members in the labor module. Column (2) replaces household size fixed effects with household (strata) fixed effects. Column (3) incorporates our full set of controls. The estimated impact of response order is similar and remains highly significant across columns.

The preferred specification in column (3) indicates that individuals are reported to be involved in 0.016 (2.2%) fewer job activities for each position further back they are in

[19]The respondents themselves can be thought of as occupying position zero, before any effects of fatigue have accrued from repetition of the labor module.

the randomized ordering. This is an economically significant effect. An individual listed last in a household with six members in the labor module (the sample median for labor module size), would expect to have more than 11% of their labor contributions lost due to response fatigue. For an individual occupying the median listing position (3), these estimates suggest expected losses just under 7%.

The remaining columns of Table 3 show robustness of these main effects to alternative codings of the outcome and treatment, as well as use of an alternative estimating model. Columns (4) and (5) change the coding of the outcome to binary indicators for reporting at least one and at least two work activities, respectively. Point estimates of effects for these two outcomes are both highly significant (p<.01) and similar to one another, -0.007 and -0.008, respectively. However, this represents a 1.3% reduction in the likelihood of recording at least one activity whereas the likelihood of listing at least two declines by 4.6% per order position. This suggests that secondary work activities may be more vulnerable to fatigue than primary activities.

Next, we check robustness of our results to a different coding of the order variable. Column (6) uses the within household percentile location of each individual (scaled so that values range from zero to one). Effects are large and highly significant suggesting average impacts of 12% when going from very first to very last in the listed order. We return to this specification later in our discussion of household size. Finally, we assess the sensitivity of the results to the choice of ordinary least squares for estimation, instead using a Poisson model in column (7) to account for the count nature of the data. Results remain negative and highly significant with effect sizes of 3.3% losses per position. Additional robustness to the sample selection criteria is shown in Appendix Table A.2.

The main results shown thus far assume that the relationship between order and reported outcomes is linear. Figure 1 explores this assumption empirically. In Figure 1 we show estimates of effects by response position, jointly estimated with the first listed member (after the respondent) omitted as the reference group. Even for the second listed individual we see a negative effect relative to the first of -0.05 ($p = 0.065$). The final coefficient, those who were tenth or later in their household listings, are reported to have 0.2

fewer distinct productive activities.[20] On average, the magnitudes of estimated negative effects increase with order. The p-value of the difference between the first and final positions is 0.06. While coefficients for individual order positions are noisily estimated and do not monotonically increase in magnitude, we consider the linear fit to be a reasonable first approximation, sufficient for the purposes of capturing the primary pattern of relevance: that response fatigue increases in order.

Next we test whether the effects of response fatigue differ by the age and gender of household members in Table 4. Column (1) repeats the preferred result from column (3) of Table 3. Column (2) shows the effects separately by gender while column (3) shows separate effects by age group. These estimates are conducted by interacting the gender or age group with response order, continuing to control for that grouping in the specification. The coefficients displayed are the estimated effects of fatigue for each sub-group, tested against the null hypothesis of no effect. Tests for the difference in effect sizes between groups are reported at the bottom of each column.

Column (2) shows that the effect of response order on the reported labor activities of men is negative, but not significantly different from zero. By contrast, the effects of fatigue on the reported labor activities of women are more than twice as large, over 3% of the mean ($p<.01$). A test for the difference between men and women is statistically significant with $p=.02$. In column (3) we examine the effects of fatigue by different age groups. Effects for the oldest group are indistinguishable from zero, whereas they are highly significant for both of the younger age groups. While the point estimates are substantially larger for 25-34 year olds than for 14-24 year olds, given their different baseline levels of labor force participation, both represent losses of approximately 3% in reported labor activities per order position for the two groups. The patterns shown are consistent with respondents valuing the labor contributions of youth and women less than those of older people and men, and indicate that the underestimation of labor activities caused by response fatigue is biased against these groups.

Finally, Table 3 suggested that over half of the jobs missed are secondary activities.

[20] We bundle those in positions ten or higher as the sample is small for these higher orderings. This final bin has approximately 5% of the sample.

We may therefore wonder if the jobs missed due to response fatigue are economically important. Appendix Table A.3 explores the effects of fatigue on different types of work, reported pay, and the extensive and intensive margins of labor supply. Though the coefficients on household farm work and wage work are similar in size, lower overall levels of wage employment mean that the proportional effect is three times larger on wage work than household farming. This under-counting of wage labor could lead to a skewed understanding of household diversification, in particular with respect to off-farm wage work in rural areas. We additionally note significant losses on the extensive margin of reporting receipt of any pay in the last week. Our results regarding the intensive margin of amount paid are noisily estimated and therefore less clear. Regarding hours worked, there is a strong negative impact on the extensive margin of doing any work in the last week, but no statistically significant impact on number of hours worked. These results suggest that response fatigue may distort our understanding of the nature of rural employment, though we do not have sufficient power to make strong claims about the economic significance of the jobs that are impacted.

3.2 Aggregate Losses from Fatigue

The estimates in the previous section show the impact of response fatigue on reported labor activities for marginal changes in roster position. However, fatigue-induced losses from one-position shifts do not capture aggregate losses that accrue through the full administration of the labor module. For each of the three models in Table 4, we use the reduced-form estimates and the original, non-randomized, household ordering to generate two counterfactuals: what the level of productive activities would have been without any losses from fatigue and what the level of productive activities would have been with fatigue if the labor module had followed the conventional practice of using the original roster ordering. With these counterfactuals, we can estimate the aggregate impacts of fatigue for the full population as well as for different sub-groups.

To predict what reported labor activities would have been in the absence of response fatigue, we use the models in Table 4. After estimating a model, we set response order to zero for everyone in the sample, and use the estimates to predict total number of activities for each individual based on their baseline characteristics. This results in a set of predicted "no fatigue" activity levels. We then create our second counterfactual by setting response order equal to each individuals' initial household roster position, as collected prior to randomization, and generate a second set of predicted activity levels. Comparing these two totals, we can then calculate the predicted aggregate losses from fatigue for the full sample as well as for different subgroups.

Table 5 presents estimates of losses from fatigue based on the three models shown in Table 4. Column (1) shows the mean number of labor activities for individuals in each sub-group indicated on the left: the full sample, males, females, and people aged 14-24, 25-34, and 35-59. The results in columns (2)-(4) use the simple model that does not allow for heterogeneity to generate predictions. Column (2) shows the predicted means without fatigue. Column (3) incorporates response fatigue with the endogenous initial household listing order. And Column (4) displays the percent losses for each group (difference between columns (2) and (3)). Columns (5)-(7) follow the same structure but use the model allowing for heterogeneity by gender while columns (8)-(10) use the model with heterogeneity by age group.

Column (4) of the first row of Table 5 shows that losses are approximately 8% in the full sample, similar to those estimated using the other two models in columns (7) and (10). The remaining rows present similar estimates for different sub-groups. Figure 2 summarizes a selection of these results with Panel A showing results by gender and Panel B by age group. Using the estimates from the simple model without heterogeneity in Panel A, we find very similar aggregate losses by gender of 7.8% for men and 8% for women. These modest differences reflect what was found in column (1) of Table 2 whereby women are listed later then men, but only very slightly, by 0.2 positions on average. However, when we incorporate heterogeneity by gender on the right side of Panel A we see sharply different impacts of fatigue by gender with women's losses estimated at over 11% and men's at just 4.3%.

We next turn to differences by age group. Even though the simple model on the left of Panel B does not incorporate differential effects of fatigue by age, we still see sharp differences in the estimated losses by age group. This is driven by the strong ordering patterns shown in Table 2 whereby young people are listed later in household rosters. The right side of Panel B introduces heterogeneity of fatigue by age group. We find that allowing for heterogeneity by age group mutes some of the impacts on the lowest age group while increasing the estimated losses for those who are 25-34. Both have losses of around 10-11%, more than five times greater than the losses of just 1.9% estimated for older adults aged 35-59. This suggests that both patterns in the listing order that place youth systematically later in their household rosters and heterogeneity in the impact of response fatigue conditional on their roster position, lead to younger individuals being disproportionately impacted and having their labor contributions under-counted.

4 Extensions and Implications

In this section we expand on several important questions for understanding and interpreting our results.

4.1 Household size

As observed in Section 2.1, average household size in our study sample is large, with an average of ten total household members and six eligible for the labor module. As such, the external validity of our findings would be reduced if response fatigue was relevant only for large households. However, our analysis shows that losses from fatigue are still significant and of meaningful magnitude for people close to the beginning of the household ordering, as noted in the earlier discussion of Figure 1. Additionally, this analysis does not account for losses from fatigue that occur between the respondent answering about themselves and the first family member they are asked about.[21] Panel A of Appendix Figure A.1 repeats this exercise restricting the sample to only households with between five and seven members in the labor module. Again, significant negative effects are identified within a few roster positions of the respondent.

[21] Because respondents were always asked about first, we can not disentangle the effect of moving from the first to the second position in the order from the difference between self and proxy reporting.

Table 6 explores household size in a different way, splitting the sample into terciles by household size. The first three columns show that order effects are negative and highly significant even for households with just 2 to 4 members in the labor module with point estimates that are, in fact, bigger in magnitude than for larger households. This does not, however, mean that the aggregate impact of fatigue is just as big (or bigger) for small households. These point estimates reflect average marginal losses from being one position further back in a household roster. With more individuals in large households, and more repetitions of the labor module, aggregate losses from fatigue are still larger for these bigger households.

This is shown in the second three columns of Table 6 by replacing the order position variable with the percentile version used in column (6) of Table 3. Reported coefficients therefore reflect the estimated difference between being randomly listed first versus last within households of different sizes. Column (4) shows that, though lacking in precision, the magnitude of estimated results for small households with 2-4 members in the labor module (excluding the respondent) have average losses of 6.4% for the final listed member. These losses climb to 11.9% for households with 5-7 members ($p<.05$) and even higher to 22.6% for households with eight or more ($p<.01$). Therefore, while the impacts of response fatigue are important for our understanding of labor participation in households of all sizes, aggregate impacts are larger in larger households. This has broad implications for work comparing labor contributions across countries, regions, or settings (such as urban and rural areas) that have different average household sizes.

Linked to household size, Table 1 showed that the sample had a relatively high prevalence of polygamy, 31%. Given that polygamous households are generally larger than non-polygamous households, this could confound our interpretation of the relationship between fatigue and household size. We may also be more broadly concerned about external validity if our results are driven by these households, where internal divisions may impede the flow of information across internal groups and exacerbate response fatigue. We check for heterogeneity between polygamous and non-polygamous households in column (2) of Table 7 and do not find significant differences. Point estimates for polygamous households are, in fact, less than half the size of those in non-polygamous households.

4.2 Proxy Reporting and Respondent Type

The use of self-reporting versus proxy response when completing the labor module is an additional possible confounder of our results. Other research has suggested that prioritizing self-reporting can reduce losses due to the difference between proxy versus self reporting (Benes and Walsh, 2018; Glewwe, 2000). It could be that this approach reduces losses from response fatigue as well. Our main specification includes an indicator for these instances of own-response as a control in our estimation and our results are also robust to dropping these observations, shown in Appendix Table A.2.[22] In Table 7 we check for heterogeneity of our main effects by whether an individual was conferred with or responded for themselves. The estimated difference in response fatigue between proxy and self-reports is close to zero, though the standard error is large, limiting our ability to draw strong conclusions.

While more heavy reliance on self-reporting could, in theory, help to mitigate the effects of response fatigue, locating and conferring with all household members is time consuming and costly and may therefore not be feasible for many studies. Many surveys adopt an approach of encouraging self-reporting wherever possible without requiring it. However, encouragement to seek self-reports can introduce its own biases related to who is and who is not available.[23] Appendix Table A.5 looks at predictors of self-reporting in our own sample as well as those of four other LSMS-style household surveys.[24] While self-reporting is especially low in our sample at 17%, the others are far from complete, ranging from 39-69%. Older individuals are significantly more likely to be reached for self-reporting in all but one of these surveys while women are substantially more likely to be reached in all five. Given that self-reporting is associated with higher reported ac-

[22]Own-response is not significantly less likely as the respondent progresses through the iterations of the labor module, shown in Table 2.

[23]The use of proxy respondents has been studied experimentally by (Bardasi et al., 2011) who find that proxy response does not affect the reporting of women's labor in Tanzania, but does lead to lower estimates of male employment.

[24]In general, LSMS field protocols emphasize that self-reporting of labor responses should be captured whenever possible.

tivity levels, partial adherence to self-reporting could unintentionally introduce greater biases by gender and age than the fatigue-induced biases it aims to solve. The extent to which heavier reliance on self-reporting does or does not improve data quality remains a topic worthy of further research.

We also explore whether response fatigue is better or worse for different types of primary respondents. In particular, factors such as gender, age, education, or patience could plausibly relate to the information respondents have about other family members or the opportunity costs of their time. In columns (3)-(6) of Table 7, we look at heterogeneity by these dimensions. Point estimates suggest that effects of fatigue may be stronger for men, more educated, younger, and less patient respondents. But differences are noisily estimated and not statistically significant. Given the difficulty in distinguishing how these groups relate to household information or opportunity costs of time and the endogeneity of who was included in the GASIP program (and therefore designated as our primary respondent), we hesitate to place excessive emphasis on these characteristics. However, we flag this as an intriguing potential mechanism deserving of further consideration and research.

4.3 Enumerator Fatigue

We also consider the extent to which these results are attributable to respondent fatigue as opposed to fatigue experienced by the enumerator. Other research has shown that enumerators can have substantial impacts on responses and data quality suggesting that enumerator fatigue could also affect resulting data (Di Maio and Fiala, 2020). Attribution of effects to respondent or enumerator would also point to different potential solutions. Thus far, we have focused on the role of the respondent because respondents ultimately control their answers and the decision of whether or not to report work activities for their household members. By contrast, enumerators are trained to follow a strict script while conducting interviews and not to lead respondents towards' any particular response. For enumerator fatigue to be driving our results, the enumerators would either need to be skipping questions and deviating from their scripts or, in some other way, be signalling that they want the respondent to under-report work activities. Additionally, to be consistent with the observed patterns of heterogeneity discussed in the analysis, enumerators would need to be differentially signalling their impatience across gender and age groups. Because the number of follow-up questions triggered by each additional work activity is the same regardless of an individual's characteristics, we would not expect enumerator fatigue to display the strong patterns of heterogeneity that were documented in Table 4.

We additionally test whether sensible patterns of enumerator fatigue can be detected in the data in Appendix Table 7. First, we test if the estimated response fatigue varies by whether or not it was the enumerator's first survey of the day, hypothesizing that enumerators would be less fatigued early on.[25] We do not find evidence of any differences in the relationship between roster position and activities. We also categorize each survey by whether it was in the first or second half of the enumerator's total number of interviews over the full course of the study. Again, we do not see any significant differential in the relationship between household roster position and reported activities by whether a survey was early or late in an enumerator's tenure.[26] Ultimately, while we cannot entirely rule out enumerator fatigue and recognize that it may still be affecting reported productive activities in other ways, our analysis suggests that it is unlikely to be driving our main results.

5 Discussion and Conclusion

Our results suggest that response fatigue can contribute to meaningful and widespread undercounting of rural labor activities. Average losses per individual are approximately 8%, or one in eleven activities. Estimated losses are especially large for youth (10-12%) and women (11%) as compared to older adults (1.8%) and men (5%). These differences create biases in the resulting data that are due both to differential vulnerability and differential exposure to fatigue for these groups. These estimates are based solely on fatigue induced by random variation in response order within a survey module. If fatigue builds

[25]The modal number of surveys performed per day is two.

[26]We also rule out that enumerators who were better rated for performance react differently to response order in results available upon request.

over the course of an entire survey, these results may present a lower bound on overall losses.

Employment opportunities for women and youth are a particular focus of researchers and policy makers. However, our results suggest that we may be systematically undercounting their contributions. Similar or even stronger ordering patterns are present in household rosters of a number of other important data sources, suggesting that biases may be present in their data as well.[27] Our analysis also indicates that secondary off-farm wage work activities are more vulnerable to undercounting due to response order; if true elsewhere, this could contribute to an understatement of the diversification of rural livelihoods. For example, Dolislager et al. (2020) provide a comprehensive accounting of youth employment across the world, and document strong sectoral and regional patterns. Given the focus on youth, the number and range of labor activities in that study may be underestimated and this underestimation may vary by sector. Where household size varies across regions, regional comparisons may also be biased by the role of response fatigue.

One possible approach to mitigating these distortions is for researchers and data collection teams to avoid the use of proxy response and instead insist on separate interviews and self-reporting. As discussed in section 4, however, this may be a prohibitive approach for less well resourced studies, will not necessarily be effective, and may introduce other biases related to individuals' availability and willingness to participate.

A second approach within the existing module structure and without cost implications is for researchers to consider their research objectives. If the primary interest is to make broad comparisons across groups or to better understand intra-household dynamics, randomization of the labor module ordering will mitigate the biases introduced by respondent fatigue. If instead, capturing aggregate household labor activities and earnings is the priority, explicitly organizing the listing in order of greatest to least economic contribution could minimize household-level losses and population wide extrapolation.

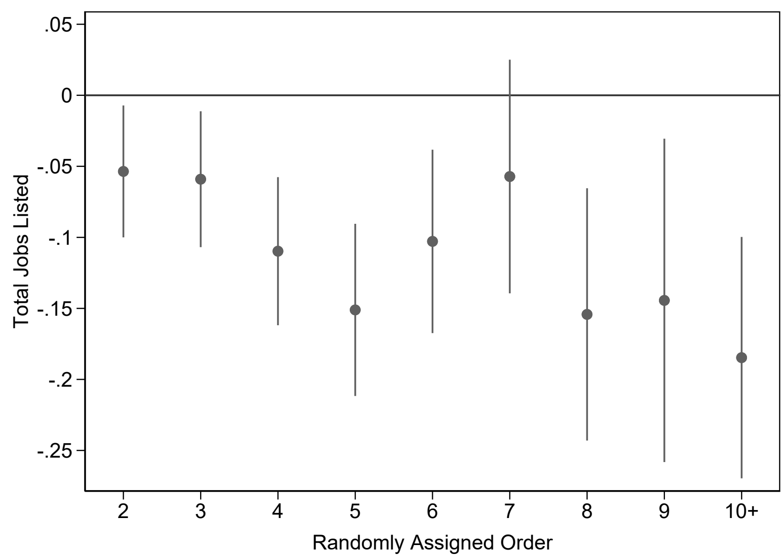
[27] Appendix Table A.4 shows that the 1993 Indonesian Family Life Survey, the 2017 Ghana GLS, the 2010 Nigerian LSMS, the 2012 Tanzanian LSMS, and the 2016 Malawi IHS all display strong ordering patterns whereby women and youth are systematically positioned later in household rosters. In these examples women are even more disadvantaged than in our own survey.

A final possibility is to explore alternative survey designs. One approach could be to document the jobs that each individual does first, before iterating over each task. Initial activity listing could be done for all household members, before asking any follow-up questions. This might avoid some of the losses from fatigue, although it may come at the cost of a more disjointed interview experience and interfere with the ability to encourage self-response from available household members. Ultimately, more methodological research is needed to understand the extent to which response fatigue impacts data quality across different contexts and topics.

Response fatigue may also extend past labor modules to other survey modules that follow repetitive structures. Topics include consumption and expenditure and agricultural production, suggesting that other important outcomes such as nutrition and household income could also be affected. Ultimately, improvement in survey design methods is key to our understanding of all aspects of peoples' livelihoods and well-being. Without a reliable evidence base, effective policy and programs aiming to improve the lives of the world's poor will suffer.

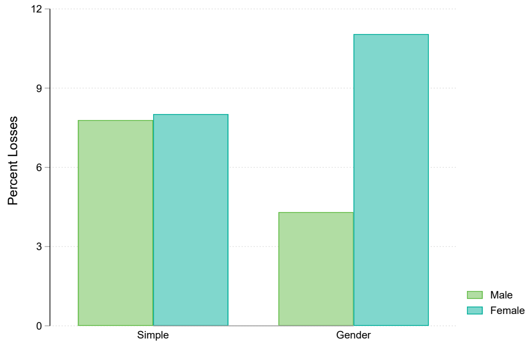
6 Main Tables and Figures

Figure 1: Effect of Randomized Order Position on Reported Total Job Activities

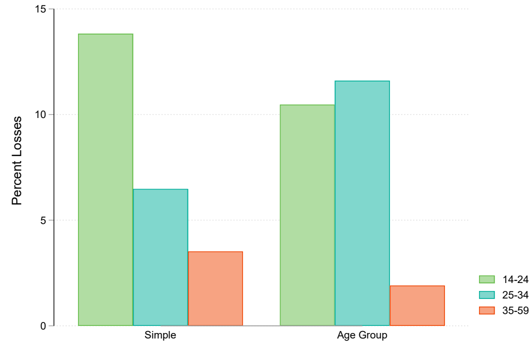


Notes: Figure shows the coefficients from jointly estimated random roster position assignments on total number of jobs listed (0-3). Estimation controls for household and relation to respondent fixed effects, gender, age group, schooling status and self-reporting.

Figure 2: Predicted Losses from Fatigue for Gender and Age Sub-Groups with Different Models



(a) Predicted Losses by Gender



(b) Predicted Losses by Age Groups

Notes: These figures show predicted aggregate losses from fatigue in reported number of labor activities for individuals by sub groups. Panel A shows differences by gender while Panel B shows differences by age group. The x-axes indicate the type of model used to predict fatigue. The "Simple" model for both panels uses the uniform fatigue factor as in the model from column (1) of Table 4. Gender heterogeneity is used in the right portion of Panel A using the model from column (2) of Table 4 and age group heterogeneity is used in the right side of Panel B as in column (3) of Table 4.

Table 1: Summary Statistics

Panel A: Household Level Characteristics	N=950	
	Mean	SD
Total Household Members	10.183	5.285
Members Included in Labor Module	5.989	2.895
Polygamous Household	0.309	0.463
Respondent: Female	0.512	0.500
Respondent: Age	42.153	11.708
Respondent: Ever School	0.325	0.469
Panel B: Household Member Characteristics	N=4252	
	Mean	SD
Female	0.534	0.499
Age	26.477	10.837
Literate	0.564	0.496
Currently in School	0.356	0.479
Self-Report	0.171	0.377
Relation to Respondent: Spouse	0.189	0.392
Relation to Respondent: Child	0.447	0.497
Relation to Respondent: Parent	0.037	0.189
Relation to Respondent: Sibling	0.123	0.328
Total Reported Job Activities	0.744	0.746
No Activities	0.436	0.496
One Activity	0.386	0.487
Two Activities	0.175	0.380
Three Activities	0.003	0.051
HH Farm Work	0.841	0.366
HH Business	0.147	0.354
Wage Work	0.253	0.435
Labor Module Order Position	3.811	2.909

Notes: Panel A shows summary statistics at the household level including household and respondent characteristics. Panel B shows characteristics of the household members who participated in the labor module (above 14 and below 60 years old), excluding the respondent. Original household roster position is among those who are at least 14 years old and who were included in the labor module. Self-Report indicates those who reported their own labor activities or were otherwise conferred with during the survey.

Table 2: Order and Balance in Analysis Sample

	(1) Listed Order	(2) Randomized Order
Female	0.212**	-0.078
	(0.101)	(0.077)
Age 25-34	-0.413***	0.015
	(0.115)	(0.118)
Age 35-59	-1.359***	0.103
	(0.150)	(0.136)
Currently in School	0.121	-0.004
	(0.107)	(0.107)
Self-Report	-0.889***	-0.164
	(0.114)	(0.151)
Rel to Resp: Child	0.654***	-0.020
	(0.159)	(0.138)
Rel to Resp: Parent	0.534**	-0.106
	(0.261)	(0.275)
Rel to Resp: Sibling	0.913***	-0.312*
	(0.184)	(0.165)
Rel to Resp: Other	1.368***	0.153
	(0.184)	(0.149)
N	4252	4252
Mean Y	3.923	3.780
Resp Gender	All	All
Households	950	950
R2	0.576	0.436

Notes: Outcome variable is an individual's within household order position. Listed order reflects individuals' roster position as recorded during initial household listing at the beginning of the survey. Randomized order reflects the implemented position used in the administration of the labor module. Self-report indicates that household member reported for self or was conferred with during labor module. Omitted age group is 14-24. Regressions include household fixed effects.

Table 3: Impact of Response Order on Reported Activities

	(1) Total	(2) Total	(3) Total	(4) One or More	(5) Two or More	(6) Total	(7) Total
Order	-0.014***	-0.015***	-0.016***	-0.007***	-0.008***	-0.089***	-0.025***
	(0.005)	(0.005)	(0.004)	(0.003)	(0.003)	(0.024)	(0.006)
Female			-0.077***	-0.081***	0.001	-0.077***	-0.130***
			(0.021)	(0.014)	(0.012)	(0.021)	(0.033)
Self-Report			0.073**	0.042**	0.034*	0.073**	0.097**
			(0.031)	(0.019)	(0.019)	(0.031)	(0.049)
Student			-0.485***	-0.372***	-0.112***	-0.485***	-1.024***
			(0.028)	(0.021)	(0.015)	(0.028)	(0.069)
Age 25-34			0.224***	0.136***	0.087***	0.222***	0.303***
			(0.031)	(0.022)	(0.017)	(0.031)	(0.045)
Age 35+			0.337***	0.211***	0.125***	0.335***	0.388***
			(0.041)	(0.026)	(0.024)	(0.041)	(0.055)
N	4252	4252	4252	4252	4252	4252	4252
Mean Y	0.744	0.744	0.744	0.564	0.178	0.744	0.744
Percent Effect	-1.919	-2.077	-2.183	-1.293	-4.609	-11.927	-3.305
Family Size FEs	Yes	No	No	No	No	No	No
Household FEs	No	Yes	Yes	Yes	Yes	Yes	Yes
Relation to Resp FEs	No	No	Yes	Yes	Yes	Yes	Yes
Model	OLS	OLS	OLS	OLS	OLS	OLS	Poisson
Order Variable	Level	Level	Level	Level	Level	Percent	Level
Households	950	950	950	950	950	950	950
R2	0.012	0.357	0.578	0.584	0.434	0.577	0.216

Notes: This table shows robustness of the estimated effects of randomized order position on the number of jobs reported for each individual. Estimates in column (3) use the preferred specification from equation (1). Order is the response order or position if this individual in the labor module sequence. Self-Report indicates that household member was conferred with during labor module. All columns code response order as a number between 1-14 except for the estimates in column (6) which code response position as the "percent" within household position calculated so that the first person in the each household is set at zero and the final person per household set at 1. “Total” is the total number of unique jobs listed for this individual. “One or More” indicates a binary outcome for having at least one job listed in the labor module. “Two or More” indicates at least two jobs listed. Self-Report indicates that household member reported for self or was conferred with during labor module. The omitted age group is 14-24.

Table 4: Impact of Response Order on Reported Activities by Gender and Age Group

Dependent Variable: Total Jobs Recorded			
	(1)	(2)	(3)
Order	-0.016***		
	(0.004)		
Order x Male		-0.009	
		(0.005)	
Order x Female		-0.023***	
		(0.005)	
Order x Age 14-24			-0.012***
			(0.004)
Order x Age 25-34			-0.031***
			(0.008)
Order x Age 35-59			-0.009
			(0.010)
N	4252	4252	4252
Group 1: Mean	.744	.738	.45
Group 2: Mean		.749	.96
Group 3: Mean			1.204
Group 1: Scaled Effect	-2.183	-1.173	-2.647
Group 2: Scaled Effect		-3.084	-3.198
Group 3: Scaled Effect			-.715
P-Val B1=B2		.02	.03
P-Val B1=B3			.754
P-Val B2=B3			.053
R2	0.578	0.578	0.578

Notes: The first column reproduces the main result from column (3) of Table 3. This table shows heterogeneity of the main effects of order position on total jobs listed. The estimates in each specification use a fully saturated treatment so that the main effect on each sub-group can be readily read from the table and tested against the null of no effect. In the statistics at the bottom of the table, group numbers indicate the relevant mean or scaled effect for the corresponding coefficients in the column above. For example, in column 3, “Group 1: Scaled Effect” suggests that youth age 14-24 have estimated losses per order position of 2.6%. “Group 5: Mean” in column 4 suggests that women between the age 25-34 have an average of 0.984 listed labor activities. The reported p-values show differences between estimated coefficients within the same model so that "P-Val B2=B3" is the p-value of the test for equality between the second and third coefficients in that model, Order x Age 25-34 and Order x Age 35-59.

Table 5: Average Losses from Fatigue with Different Models for Different Subgroups

		Model Heterogeneity								
		Model 1: No Heterogeneity			Model 2: Gender			Model 3: Age Group		
Sub-Sample	(1) Sample Mean	(2) No Fatigue	(3) Fatigue	(4) Difference	(5) No Fatigue	(6) Fatigue	(7) Difference	(8) No Fatigue	(9) Fatigue	(10) Difference
All	0.744	0.806	0.742	-7.912%	0.806	0.741	-8.043%	0.805	0.741	-7.900%
	(0.746)	(0.565)	(0.574)		(0.566)	(0.574)		(0.571)	(0.572)	
Male	0.738	0.799	0.737	-7.791%	0.771	0.737	-4.305%	0.797	.0736	-7.632%
	(0.726)	(0.561)	(0.570)		(0.561)	(0.566)		(0.564)	(0.569)	
Female	0.749	0.811	0.746	-8.015%	0.837	0.745	-11.047%	0.811	0.745	-8.130%
	(0.763)	(0.568)	(0.577)		(0.568)	(0.581)		(0.576)	(0.576)	
Age 14-24	0.450	0.511	0.440	-13.824%	0.510	0.439	-13.843%	0.495	0.443	-10.474%
	(0.649)	(0.474)	(0.477)		(0.473)	(0.477)		(0.474)	(0.476)	
Age 25-34	0.960	1.024	0.957	-6.482%	1.027	0.957	-6.771%	1.081	0.955	-11.603%
	(0.737)	(0.472)	(0.479)		(0.476)	(0.479)		(0.473)	(0.488)	
Age 35-59	1.204	1.265	1.220	-3.524%	1.267	1.221	-3.641%	1.236	1.213	-1.909%
	(0.652)	(0.423)	(0.430)		(0.421)	(0.429)		(0.424)	(0.427)	

Notes: Each row presents means (and standard deviations in parentheses) for the sample listed to the left of the row. The values in column (1) are numbers of jobs as reported in the raw data. Columns (2)-(4) use the simple model of fatigue without heterogeneity, as estimated in column (1) of Table 4. Column (2) removes predicted fatigue by setting $order_i = 0$ for all individuals then calculating the mean of these predicted "No Fatigue" values. Column (3) reimposes predicted fatigue as if individuals had been asked about in the original, non-randomized, household listing order. Column (4) calculates the proportion losses from fatigue, the difference between columns (2) and (3) divided by the mean in column (2). This approach for columns (2)-(4) is then repeated in columns (5)-(7) and again in columns (8)-(10) by using predicted no fatigue and fatigue means based off of models that allow for different dimenions of heterogeneity. Columns (5)-(7) allow for heterogeneity by gender as estimated in column (2) of Table 4 while Columns (8)-(10) allow for heterogeneity by age group as estimated in column (3) of Table 4.

Table 6: Fatigue and Respondent Order by Household Size

	Number (1-14)			**Percentile (0-1)**		
	(1)	(2)	(3)	(4)	(5)	(6)
Order	-0.035**	-0.016*	-0.015***	-0.048	-0.089**	-0.168***
	(0.016)	(0.009)	(0.005)	(0.035)	(0.042)	(0.050)
N	1414	1551	1287	1414	1551	1287
Mean Y	0.786	0.740	0.703	0.786	0.740	0.703
Percent Effect	-4.645	-2.170	-1.997	-6.427	-11.930	-22.575
Family Size	2-4	5-7	8+	2-4	5-7	8+
Households	509	302	139	509	302	139
R2	0.653	0.563	0.520	0.652	0.563	0.521

Notes: Dependent Variable: Total Jobs Recorded. Order is the response order or position if this individual in the labor module sequence. Columns (1)-(3) use the order number that the household member was asked about in the labor module from 1-14. Columns (4)-(6) use the within household percentile position in place of the order number ranging from zero (first position) to one (last position).

Table 7: Additional Fatigue Effect Heterogeneity

	Dependent Variable: Labor Activities Reported							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Order	-0.023***	-0.016***	-0.020***	-0.025***	-0.012**	-0.022***	-0.015**	-0.018***
	(0.006)	(0.005)	(0.005)	(0.009)	(0.005)	(0.007)	(0.006)	(0.006)
Order x Household Polygamous	0.012							
	(0.008)							
Order x Member Self-Report		-0.001						
		(0.013)						
Order x Respondent Female			0.009					
			(0.009)					
Order x Respondent Older				0.013				
				(0.010)				
Order x Respondent Educated					-0.013			
					(0.010)			
Order x Respondent Patient						0.009		
						(0.009)		
Order x Enum Not First of Day							-0.003	
							(0.008)	
Order x Enum Second Half								0.004
								(0.008)
N	4252	4252	4252	4252	4252	4252	4252	4252
Mean Y	0.744	0.744	0.744	0.744	0.744	0.744	0.744	0.744
R2	0.578	0.578	0.578	0.578	0.578	0.578	0.578	0.578

Notes: This table shows heterogeneity of fatigue effects along a number of different dimensions by interacting the main treatment variable, "Order", with different characteristics. Self-report varies at the individual level (and is controlled for in the specification) whereas all of the others are constant within household and thus absorbed by the household fixed effects. Enum is a characteristic of the enumerator, whether the interview was second or third (not first) of the day and whether it was in the second half of all of that enumerators work for the duration of the entire project. Respondent indicates characteristics of the household's designated respondent: gender, any education, and above median age among respondents.

References

Ambler, Kate, Alan de Brauw, and Mike Murphy (2020) "Increasing the adoption of conservation agriculture: A framed field experiment in Northern Ghana," URL: `https://ebrary.ifpri.org/digital/collection/p15738coll2/id/133739`.

Arthi, Vellore, Kathleen Beegle, Joachim De Weerdt, and Amparo Palacios-Lopez (2016) *Not your average job: Measuring farm labor in Tanzania*: The World Bank.

Asfaw, Solomon, Antonio Scognamillo, Gloria Di Caprera, Nicholas Sitko, and Adriana Ignaciuk (2019) "Heterogeneous impact of livelihood diversification on household welfare: Cross-country evidence from Sub-Saharan Africa," *World Development*, Vol. 117, pp. 278–295.

Bardasi, Elena, Kathleen Beegle, Andrew Dillon, and Pieter Serneels (2011) "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania," *The World Bank Economic Review*, Vol. 25, pp. 418–447.

Beaman, Lori and Andrew Dillon (2012) "Do household definitions matter in survey design? Results from a randomized survey experiment in Mali," *Journal of Development Economics*, Vol. 98, pp. 124–135.

Beegle, Kathleen, Calogero Carletto, and Kristen Himelein (2012a) "Reliability of recall in agricultural data," *Journal of Development Economics*, Vol. 98, pp. 34–41.

Beegle, Kathleen, Joachim De Weerdt, Jed Friedman, and John Gibson (2012b) "Methods of Household Consumption Measurement through Surveys: Experimental Results from Tanzania," *Journal of Development Economics*, Vol. 98, pp. 3–18.

Benes, Elisa and Kieran Walsh (2018) "Measuring employment in Labour Force Surveys: main findings from the ILO LFS pilot studies," *ILO*.

Betcherman, Gordon and Themrise Khan (2018) "Jobs for Africa's expanding youth cohort: a stocktaking of employment prospects and policy interventions," *IZA Journal of Development and Migration*, Vol. 8, p. 13.

Bradley, Mark and Andrew Daly (1994) "Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data," *Transportation*, Vol. 21, pp. 167–184.

Bridges, Sarah, David Lawson, and Sharifa Begum (2011) "Labour market outcomes in Bangladesh: The role of poverty and gender norms," *The European Journal of Development Research*, Vol. 23, pp. 459–487.

Van den Broeck, Goedele and Talip Kilic (2019) "Dynamics of off-farm employment in Sub-Saharan Africa: A gender perspective," *World Development*, Vol. 119, pp. 81–99.

Caeyers, Bet, Neil Chalmers, and Joachim De Weerdt (2012) "Improving consumption measurement and other survey data through CAPI: Evidence from a randomized experiment," *Journal of Development Economics*, Vol. 98, pp. 19–33.

Comblon, Virginie, Anne-Sophie Robilliard et al. (2015) "Are female employment statistics more sensitive than male ones to questionnaire design? Evidence from Cameroon, Mali and Senegal,"Technical report.

Das, Jishnu, Jeffrey Hammer, and Carolina Sánchez-Paramo (2012) "The impact of recall periods on reported morbidity and health seeking behavior," *Journal of Development Economics*, Vol. 98, pp. 76–88.

Davis, Benjamin, Paul Winters, Gero Carletto, Katia Covarrubias, Esteban J Quiñones, Alberto Zezza, Kostas Stamoulis, Carlo Azzarri, and Stefania DiGiuseppe (2010) "A cross-country comparison of rural income generating activities," *World development*, Vol. 38, pp. 48–63.

Deininger, Klaus, Calogero Carletto, Sara Savastano, and James Muwonge (2012) "Can diaries help in improving agricultural production statistics? Evidence from Uganda," *Journal of Development Economics*, pp. 42–50.

Desiere, Sam and Valentina Costa (2019) *Employment Data in Household Surveys: Taking Stock, Looking Ahead*: The World Bank.

Di Maio, Michele and Nathan Fiala (2020) "Be Wary of Those Who Ask: A Randomized Experiment on the Size and Determinants of the Enumerator Effect," *The World Bank Economic Review*, Vol. 34, pp. 654–669.

Dillon, Andrew, Elena Bardasi, Kathleen Beegle, and Pieter Serneels (2012) "Explaining Variation in Child Labor Statistics," *Journal of Development Economics*, Vol. 98, pp. 136–147.

Djurfeldt, Agnes Andersson (2013) "African re-agrarianization? Accumulation or pro-poor agricultural growth?" *World Development*, Vol. 41, pp. 217–231.

Dolislager, Michael, Thomas Reardon, Aslihan Arslan, Louise Fox, Saweda Liverpool-Tasie, Christine Sauer, and David L Tschirley (2020) "Youth and adult agrifood system employment in developing regions: Rural (peri-urban to hinterland) vs. urban," *The Journal of Development Studies*, pp. 1–23.

Dzanku, Fred Mawunyo (2020) "Poverty Reduction and Economic Livelihood Mobility in Rural Sub-Saharan Africa," *Journal of International Development*.

Egleston, Brian L, Suzanne M Miller, and Neal J Meropol (2011) "The impact of misclassification due to survey response fatigue on estimation and identifiability of treatment effects," *Statistics in medicine*, Vol. 30, pp. 3560–3572.

Ellis, Frank (1998) "Household strategies and rural livelihood diversification," *Journal of Development Studies*, Vol. 35, pp. 1–38.

Ellis, Frank and H Ade Freeman (2004) "Rural livelihoods and poverty reduction strategies in four African countries," *Journal of Development Studies*, Vol. 40, pp. 1–30.

Fafchamps, Marcel, David McKenzie, Simon Quinn, and Christopher Woodruff (2012) "Using PDA consistency checks to increase the precision of profits and sales measurement in panels," *Journal of Development Economics*, Vol. 98, pp. 51–57.

Fox, Louise and Obert Pimhidzai (2013) *Different Dreams, Same Bed: Collecting, Using, and Interpreting Employment Statistics in Sub-Saharan Africa-The Case of Uganda*: The World Bank.

Gaddis, Isis, Gbemisola Oseni, Amparo Palacios-Lopez, and Janneke Pieters (2020) "Measuring Farm Labor: Survey Experimental Evidence from Ghana," *The World Bank Economic Review*, DOI: `http://dx.doi.org/10.1093/wber/lhaa012`.

Galdo, Jose, Ana C Dammert, and Degnet Abebaw (2020) "Gender Bias in Agricultural Child Labor," *Working Paper*.

Galesic, Mirta and Michael Bosnjak (2009) "Effects of questionnaire length on participation and indicators of response quality in a web survey," *Public opinion quarterly*, Vol. 73, pp. 349–360.

Glewwe, Paul (2000) "Household Roster," in Margaret Grosh and Paul Glewwe eds. *Designing household survey questionnaires for developing countries*: The World Bank, Chap. 6, pp. 135–141.

Grosh, Margaret, Paul Glewwe et al. (2000) *Designing Household Survey Questionnaires for Developing Countries*: The World Bank.

Haggblade, Steven, Peter Hazell, and Thomas Reardon (2010) "The rural non-farm economy: Prospects for growth and poverty reduction," *World Development*, Vol. 38, pp. 1429–1441.

Hart, Timothy C, Callie Marie Rennison, and Chris Gibson (2005) "Revisiting respondent "fatigue bias" in the National Crime Victimization Survey," *Journal of Quantitative Criminology*, Vol. 21, pp. 345–363.

Heath, Rachel, Ghazala Mansuri, Bob Rijkers, William Hutchins Seitz, and Dhiraj Sharma (2020) "Measuring Employment: Experimental Evidence from Urban Ghana," *World Bank Policy Research Working Paper*.

Hess, Stephane, David A Hensher, and Andrew Daly (2012) "Not bored yet–revisiting respondent fatigue in stated choice experiments," *Transportation research part A: policy and practice*, Vol. 46, pp. 626–644.

Himanshu, Peter Lanjouw, Rinku Murgai, and Nicholas Stern (2013) *Non-farm diversification, poverty, economic mobility and income inequality: A case study in village India*: The World Bank.

Holbrook, Allyson L., Jon A. Krosnick, David Moore, and Roger Tourangeau (2007) "Response Order Effects in Dichotomous Categorical Questions Presented Orally: The Impact of Question and Respondent Attributes," *Public Opinion Quarterly*, Vol. 71, pp. 325–348.

Imai, Katsushi S, Raghav Gaiha, and Ganesh Thapa (2015) "Does non-farm sector employment reduce rural poverty and vulnerability? Evidence from Vietnam and India," *Journal of Asian Economics*, Vol. 36, pp. 47–61.

Klasen, Stephan and Francesca Lamanna (2009) "The impact of gender inequality in education and employment on economic growth: new evidence for a panel of countries," *Feminist Economics*, Vol. 15, pp. 91–132.

Krumbiegel, Katharina, Miet Maertens, and Meike Wollni (2020) "Can employment empower women? Female workers in the pineapple sector in Ghana," *Journal of Rural Studies*.

Laajaj, Rachid and Karen Macours (forthcoming) "Measuring Skills in Developing Countries," *The Journal of Human Resources*.

Langsten, Ray and Rania Salen (2008) "Two approaches to measuring women's work in developing countries: A comparison of survey data from Egypt," *Population and Development Review*, Vol. 34, pp. 283–305.

Martin, Elizabeth and Anne E Polivka (1995) "Diagnostics for redesigning survey questionnaires: Measuring work in the Current Population Survey," *Public Opinion Quarterly*, Vol. 59, pp. 547–567.

McKenzie, David and Mark Rosenzweig (2012) "Preface for symposium on measurement and survey design," *Journal of Development Economics*, Vol. 98, pp. 1–2.

Roberts, Caroline, Gillian Eva, Nick Allum, and Peter Lynn (2010) "Data quality in telephone surveys and the effect of questionnaire length: A cross-national experiment,"Technical report, ISER Working Paper Series.

Rolstad, Sindre, John Adler, and Anna Rydén (2011) "Response burden and questionnaire length: is shorter better? A review and meta-analysis," *Value in Health*, Vol. 14, pp. 1101–1108.

Schaffner, Julie Anderson (2000) "Employment," in Margaret Grosh and Paul Glewwe eds. *Designing household survey questionnaires for developing countries*: The World Bank, Chap. 9, pp. 135–141.

Serneels, Pieter, Kathleen Beegle, and Andrew Dillon (2016) *Do Returns to Education Depend on How and Who You Ask?*, Policy Research Working Papers: The World Bank, DOI: `http://dx.doi.org/10.1596/1813-9450-7747`.

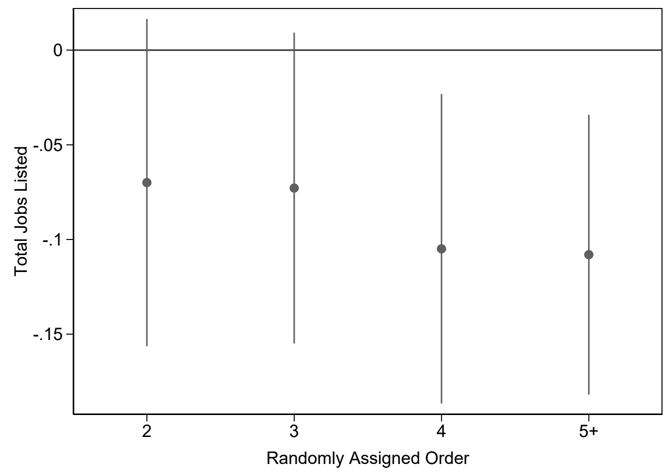
Sharp, Laure M and Joanne Frankel (1983) "Respondent burden: A test of some common assumptions," *Public Opinion Quarterly*, Vol. 47, pp. 36–53.

Yeboah, Felix Kwame and Thomas S Jayne (2018) "Africa's evolving employment trends," *The Journal of Development Studies*, Vol. 54, pp. 803–832.

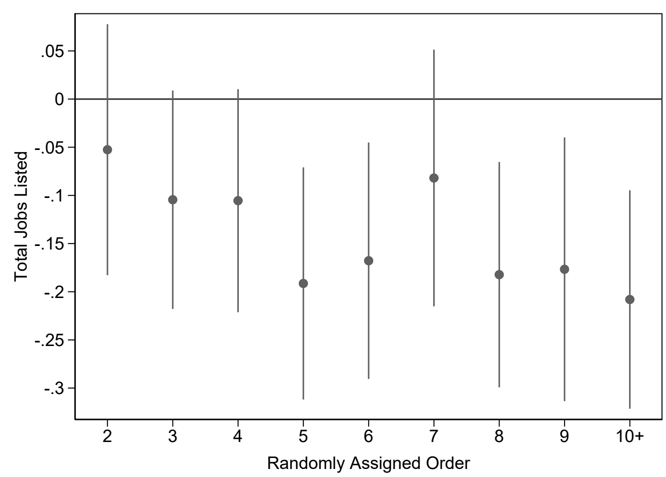
7 Appendix

Figure A.1: Ordering Effects on Subsets of Household Size

(a) Households with 5-7 People in Labor Module



(b) Households with 8+ People in Labor Module



Notes: Individual effects by labor module order position for household size subgroups.

Table A.1: Summary Statistics for Raw and Analysis Samples

Sample:	**Full**		**3+ Eligible**		**3+ Eligible No Seniors**	
Panel A: Household Level Characteristics	N=1106		N=988		N=950	
	Mean	SD	Mean	SD	Mean	SD
Total Household Members	9.476	5.234	10.012	5.261	10.183	5.285
Members Included in Labor Module	5.462	2.987	5.876	2.895	5.989	2.895
Polygamous Household	0.276	0.447	0.301	0.459	0.309	0.463
Respondent: Female	0.517	0.500	0.515	0.500	0.512	0.500
Respondent: Age	41.188	11.701	41.893	11.740	42.153	11.708
Respondent: Ever School	0.349	0.477	0.332	0.471	0.325	0.469
Sample:	**Full**		**3+ Eligible**		**3+ Eligible No Seniors**	
Panel B: Household Member Characteristics	N=4935		N=4817		N=4252	
	Mean	SD	Mean	SD	Mean	SD
Female	0.541	0.498	0.542	0.498	0.534	0.499
Age	30.998	16.227	30.940	16.315	26.477	10.837
Literate	0.502	0.500	0.506	0.500	0.564	0.496
Currently in School	0.311	0.463	0.317	0.465	0.356	0.479
Conferred on Labor	0.182	0.386	0.176	0.381	0.171	0.377
Relation to Respondent: Spouse	0.206	0.404	0.191	0.393	0.189	0.392
Relation to Respondent: Child	0.392	0.488	0.400	0.490	0.447	0.497
Relation to Respondent:Parent	0.097	0.296	0.098	0.298	0.037	0.189
Relation to Respondent: Sibling	0.111	0.315	0.114	0.318	0.123	0.328
Total Reported Job Activities	0.745	0.749	0.734	0.746	0.744	0.746
No Activities	0.437	0.496	0.444	0.497	0.436	0.496
One Activity	0.384	0.486	0.381	0.486	0.386	0.487
Two Activities	0.176	0.381	0.173	0.378	0.175	0.380
Three Activities	0.003	0.055	0.002	0.050	0.003	0.051
Household Farm Work among Working	0.850	0.357	0.846	0.361	0.841	0.366
Household Business among Working	0.145	0.352	0.146	0.353	0.147	0.354
Wage Work among Working	0.247	0.431	0.249	0.432	0.253	0.435
Labor Module Order Position	3.730	2.897	3.797	2.901	3.811	2.909

Notes: Column headers indicate sample. The analysis sample used in the analysis is reproduced for comparison in the final two columns. The first pair of columns characterize the full set of households and individuals in the raw data. The second two remove households with less than three individuals who participated in the labor module. This is because respondents are removed from the analysis of responses order and fatigue and the inclusion of household fixed effects requires a minimum of two additional people in a household who participated in the labor module in order to estimate. Finally, we remove seniors in the final two columns which both drops individuals above 60+ as well as households who, without seniors do not have sufficient members to estimate in the analysis.

Table A.2: Robustness to Sample Definition

	Dependent Variable: Labor Activities Recorded					
	(1)	(2)	(3)	(4)	(5)	(6)
Order (1-14)	-0.017***	-0.017***	-0.019***	-0.016***	-0.016***	-0.018***
	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.006)
N	4817	3879	3122	4252	4582	1938
Excluded Groups	Full	Self-Report	Students	Seniors: 60+	Incapacitated	All Restrictions
Mean Y	0.734	0.711	0.935	0.744	0.767	0.977
Percent Effect	-2.222	-2.216	-2.488	-2.179	-2.171	-2.400
Households	988	874	793	950	968	554
R2	0.523	0.534	0.433	0.578	0.561	0.507

Notes: All regressions include household, age group, and relation to respondent fixed effects as well as controls for gender, student, and being conferred with in the labor module. “Full” is the full set of individuals from the raw data. “Excluded groups” indicates sample exclusion criteria. Incapacitated are people revealed to be too old or unfit to work.

Table A.3: Work Types, Labor Supply, and Pay

	Type of Work			Pay Last Week			Hours Last Week		
	(1) HH Farm	(2) HH Bus	(3) Wage Work	(4) Any Pay	(5) IH(Pay)	(6) Log(Pay)	(7) Any Hours	(8) Hours	(9) Log(Hours)
Order (1-14)	-0.006**	-0.001	-0.005**	-0.006***	-0.035***	-0.020	-0.007**	-0.130	0.008
	(0.003)	(0.002)	(0.002)	(0.002)	(0.010)	(0.027)	(0.003)	(0.106)	(0.008)
N	4252	4252	4252	4252	4219	446	4252	4219	1112
Mean Y	0.474	0.083	0.143	0.162	0.698	3.771	0.341	10.600	3.159
Percent Effect	-1.222	-1.557	-3.775	-3.957	-	-	-1.920	-1.228	-
All	All	All	All	All	All	All	All	All	All
R2	0.571	0.389	0.388	0.438	0.442	0.734	0.489	0.478	0.753

Notes: This table looks at impacs of fatigue on additional labor related outcomes. Percent effects are not reported for outcomes converted with the inverse hyperbolic sine transformation (IH) or logs, as indicated in the column headers, as these percents can be read directly from the regression results.

Table A.4: Predictors of Household Roster Listing Order in Other Data Sources

	(1) Ghana GLS 7	(2) Nigeria LSMS 1	(3) Tanzania LSMS 3	(4) Malawi IHS 4	(5) Indonesia IFLS 1
Female	0.499***	0.792***	0.451***	0.377***	0.355***
	(0.017)	(0.030)	(0.026)	(0.014)	(0.020)
Age 25-34	-1.098***	-1.036***	-0.934***	-0.927***	-1.004***
	(0.035)	(0.057)	(0.051)	(0.033)	(0.038)
Age 35-59	-2.039***	-2.252***	-2.307***	-1.778***	-2.032***
	(0.025)	(0.048)	(0.033)	(0.024)	(0.026)
N	19784	12407	9563	13767	13513
Mean Y	2.823	3.195	2.939	2.551	2.758
R2	0.634	0.606	0.632	0.730	0.633

Notes: Dependent variable is household roster position among individuals age 14 and up. Omitted age group is youth age 14-24. Column headers indicate survey data sources. Regressions include household and relation to household head fixed effects. Regressions clustered at the household level.

Table A.5: Predictors of Self Reporting in Labor Module

	Dependent Variable: Self-Reporting in Labor Module				
	(1)	(2)	(3)	(4)	(5)
	GASIP	Ghana GLSS 7	Nigeria GHS 1	Tanzania NPS 3	Malawi IHS 4
Female	0.022*	0.037***	-0.128***	0.135***	0.143***
	(0.013)	(0.007)	(0.009)	(0.009)	(0.009)
Age 25-34	0.059***	0.117***	0.098***	0.060***	0.285***
	(0.016)	(0.012)	(0.014)	(0.015)	(0.017)
Age 35-59	0.101***	0.267***	0.284***	0.214***	0.377***
	(0.019)	(0.009)	(0.012)	(0.012)	(0.013)
N	4252	19265	12304	9436	13767
Mean Y	0.171	0.609	0.536	0.690	0.390
R2	0.360	0.484	0.547	0.371	0.471

Notes: This table shows predictors of self-reporting in the labor module. Column headers indicate the data source. GASIP is the original data collected for the project whereas the other four are all LSMS surveys. Note that the GASIP data omits the primary respondent (who always self-reported their labor contributions) whereas the other data sources do not designate a primary respondent for the labor module even while incidence of proxy reporting are high ranging from 30-60% (equal to one minus the mean level of self-reporting). All regressions include household fixed effects and control for being in-school and are clustered at the household level. GLSS=Ghana Living Standards Survey. GHS=General Household Survey. IHS=Integrated Household Survey. The number indicates the survey round.