Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania

Elena Bardasi, Kathleen Beegle, Andrew Dillon, and Pieter Serneels

Labor market statistics are critical for assessing and understanding economic development. However, widespread variation exists in how labor statistics are collected in household surveys. This paper analyzes the effects of alternative survey design on employment statistics by implementing a randomized survey experiment in Tanzania. Two features of the survey design are assessed – the level of detail of the employment questions and the type of respondent. It turns out that both features have relevant and statistically significant effects on employment statistics. Using a short labor module without screening questions induces many individuals to adopt a broad definition of employment, incorrectly including domestic duties. But after reclassifying those in domestic work as 'not working' in order to obtain the correct ILO classification, the short module turns out to generate lower female employment rates, higher working hours for both men and women who are employed, and lower rates of wage employment than the detailed module. Response by proxy rather than self-report has no effect on female labor statistics but yields substantially lower male employment rates, mostly due to underreporting of agricultural activity. The large impacts of proxy responses on male employment rates are attenuated when proxy informants are spouses and individuals with some schooling. JEL CODES: J21, C83, C93.

I. INTRODUCTION

Labor market statistics are critical for assessing how an economy functions, but they may be sensitive to the survey method by which they are collected. This paper provides evidence of the impact of the respondent type (self-reporting vs. proxy informant) and the level of detail of the labor module on labor force participation, hours worked, earnings, sector distribution, and employment status from a randomized survey experiment in a low income country. The findings confirm that labor market statistics are indeed sensitive to the survey method. In particular, male employment is especially sensitive to the selection of the informant, while female employment varies in relation to the inclusion of screening questions at the beginning of the employment module.

The experiment is carried out in a low-income country – Tanzania – and contributes to the scarce literature on survey methodology in developing countries. A wealth of evidence exists on the quality and reliability of labor statistics in household surveys, coming largely from the United States (see Bound et al., 2001, for a thorough review). However, few studies in low-income countries have attempted to rigorously measure the effect of specific features of the survey on the employment statistics it generates. These effects may vary across groups in the population – for example, employment statistics for women and children may be particularly sensitive to the survey method.[1] When the wording of the employment questions emphasizes the main activity, this may lead to underestimating the number of economically active women because of the large female presence among unpaid agricultural and family workers (Dixon-Mueller and Anker, 1988). Child and teenage work may be similarly underreported.

One way to explore whether alternative survey designs impact the labor statistics they produce would be to examine data from two surveys of different design but covering the same country and time period. This is the approach of Guarcello et al. (2009) in examining child labor. However, a review of national surveys in low-income countries yielded few relevant examples of surveys measuring employment and carried out in the same country at sufficiently close moments in time. This reflects the fact that for most low-income countries, the national surveys are either not annual or, if they are, they are topic specific (such as the Labor Force Survey, followed the next year by the Demographic Health Survey, followed by the Household Budget Survey). For Tanzania, the Integrated Labour Force Survey (ILFS) 2000/01 reports labor force participation rates of 90.6 percent for men and 89.5 percent for women (NBS, 2003), while the Household Budget Survey (HBS) reports 91.1 and 82.4 percent,

---

1. Guarcello et al. (2009) review discrepancies in child labor statistics across surveys in several low-income countries. Dillon et al. (2010) study the effect of survey design on child labor statistics using data from the same experiment as used for this paper.

respectively, for the same year (NBS, 2002). The large difference in labor force participation of women between two nationwide surveys that refer to the same year may reflect genuine differences in samples or in the timing of the survey, or may reflect differences caused by the use of distinct survey instruments.[2]

Another interesting example is offered by the Malawi Integrated Household Survey 2004/05. Although not designed as an experiment, this survey included questions on labor both in the household roster (main activity of members reported by the head) and in a module listing several non-mutually exclusive activities (supposed to be self-reported). The results show that the main activity question, answered by the head of the household, understates the percentage of individuals in farming and paid employment and overstates the fraction of inactive people (both men and women), compared with the activity-specific set of questions, answered by each individual.[3] While in this case the effects due to the interviewer and the timing of the survey are controlled for, it is not possible to attribute these differences to either the type of question or the type of respondent; only a randomized design would allow for this.

Assessing the internal validity of survey measures, although common in the psychological sciences, remains scarce in economics. In this paper, the focus is on two key survey characteristics: the level of detail of the employment questions and the informant type. There is an experimental setting to estimate the impact of each feature. The results show that there are significant differences in labor force participation, type of economic activity, and hours worked across survey designs. Using a labor module with no screening questions generates lower female labor force participation once domestic work is reclassified as no work to be in line with standard definitions, and higher average working hours

2. The question on employment indeed differs between the HBS and the ILFS. The HBS uses one single question to collect information about participation in employment and type of activity (or category of inactivity): "During the last 7 days what was your main activity?" The individual can choose among eleven categories of employment (farming/livestock keeping; fishing; mining; tourism; government employee; parastatal employee; NGO/religious organization employee; private or other employee; self-employed with employee; self-employed without employee; unpaid family helper in non-agriculture business), unemployment, and seven categories of inactivity (no activity; household chores; student; not active: retired; not active: sick; not active: disabled; not active: other). The question is repeated for the second activity. All categories are listed without any explicit distinction between employment and non-employment categories; quite interestingly, four categories are explicitly labeled as 'not active', when in fact there are another three categories of inactivity. In the ILFS, one first question asks about the usual activities during the last 12 months, to be chosen among a list of 43 economic activities, with options as detailed as "agriculture: cash crop: cotton", or "construction: farm buildings or fences". Multiple answers are allowed. Information about current activity is also recorded, with reference to the same list of economic activities used to identify the usual activity. Household duties and other categories of inactivity are explicitly labeled as such and offered as an option in a later question ("What was your main activity when you were not doing economic activity and not available for work during that period?"). By providing a detailed list of economic activities clearly defined as 'work' the ILFS explicitly defines what employment is, while the HBS does not. It is not possible, however, to determine that this is the source of the discrepancy between the employment statistics for women because other survey elements differ as well (the number of categories, the sequence of questions, etc.).

3. Detailed results are available from the authors upon request.

for both men and women. Response by proxy rather than self-report yields substantially lower male labor force participation, lower male working hours, and lower employment in agriculture for men. This indicates that the survey design matters to measuring labor outcomes and, moreover, that comparisons across surveys with different design can be compromised by these differences.

The structure of the paper is as follows. In the next section, key findings from studies from high-income countries and developing countries are discussed. In Section 3, the experimental design is described. Section 4 provides a description of the data, while Section 5 presents the results. Section 6 concludes.

II. Background and Literature

Bias in statistics from surveys can arise from several sources. Besides sampling error (related to sample size) and poor representativeness of the sample (due to non-response bias, under-coverage of certain groups of the population, or respondent self-selection), an important source of bias in surveys is measurement error. Measurement error – the difference between the value of a characteristic reported in the survey and the ("true" and unknown) value sought by the researcher – is related to the data-collection process. Its main sources are the questionnaire (question selection, sequencing, and wording), the type of informant, the data-collection method, and the interviewer. One should be concerned about measurement error because it may bias both survey statistics as well as estimates of relationships between measures of employment and other variables – hence the importance of understanding how different survey methods may impact the accuracy of the data collected.

The sources of measurement error have been studied mostly in the context of developed countries (Bound, Brown, Mathiowetz, 2001; Biemer et al., 1991). A recent review of the main issues concerning survey design in developing and transition countries relies almost entirely on research in developed countries when discussing measurement error (Kasprzyk, 2005). As a matter of fact, however, very little research exists for developing countries and it is not clear how relevant the methodological literature from high-income countries is for low-income countries. One main difference is represented by the variables of interest – even when focusing on the restricted area of employment statistics. In developed economies, for example, a lot of emphasis has been placed on the correct measurement of unemployment, especially in relation to the use of panel data for the measurement of unemployment duration and the transitions to and from unemployment.[4] In low-income countries, however, the concept of unemployment, as defined by the International Labour Organization (ILO) and understood in developed societies, seems less relevant;

---

4. Some examples include Poterba and Summers (1986, 1995), Sinclair and Gastwirth (1998), and Singh and Rao (1995).

it is rather the concept of employment (and its quality and intensity) that is at the same time important and elusive for the researcher. This explains why employment is a main variable of interest, alongside earnings and hours of work.

The experiment focuses on two sources of measurement error, specifically on the effects of (i) using detailed probing questions vs. a single, shorter question, and (ii) using proxy informants instead of self-reports. In the brief review of the literature on these two sources of measurement error, references are mostly to the studies that are relevant for this analysis – those papers analyzing impacts on employment and possibly adopting an experimental framework.

The specific wording and style of employment questions are posited to have a large influence on labor statistics. This may be particularly relevant in a setting where a significant proportion of individuals are employed in household-owned enterprises or home production and are not directly remunerated in the form of a salary or wage. For example, the standard question "Did you *work* in the last 7 days?" is hypothesized to systematically undercount persons who work in household enterprise activities without direct wage payments (e.g., unpaid family workers), who may have difficulties in identifying themselves as 'working'. Likewise, employment questions that only cover the past 7 days may produce incorrect statistics on employment participation in settings where employment is highly seasonal or where a significant proportion of workers are casual laborers.

A number of studies have focused on the style of different questions (open vs. closed questions, positive vs. negative statements, etc.) and the effects of their placement in the survey questionnaire (see the review in Kalton and Schuman, 1982). Mostly they have confirmed that question-wording effects are important, although the direction of these effects is often unpredictable. Studies have been carried out in the context of the revision of the employment questions in the U.S. Current Population Survey (CPS) to investigate the concern that irregular, unpaid, and marginal activities may be underreported partly because people do not think of themselves as working. In the Respondent Debriefing Study, respondents were asked to classify hypothetical situations ("vignettes") in terms of "work," "job," "business," and so on. Generally, the majority of respondents were able to classify the situations consistently with definitions of the CPS. However, for each vignette, large minorities of respondents gave incorrect answers – for example, 38 percent of the respondents included non-work activities under the "work" classification (Campanelli, Rothgeb, and Martin, 1989).[5] An experiment carried out in 1991 to assess the revision of the CPS questionnaire using vignettes and direct screening questions for unreported work confirmed that questionnaire wording

---

5. Esposito et al. (1991) discuss methodological tools used to obtain diagnostic information to evaluate the effect of questionnaire revisions on reporting of work activities, including hypothetical vignettes and direct screening questions.

and sequence of questions affect the respondent's interpretation of work and, therefore, the employment statistics (Martin and Polivka, 1995). Moreover, the use of direct screening questions was found particularly useful to detect underreporting of work done in connection with the household business or farm, as well as underreporting of teenage employment.

The 1991 CPS study noted above also pointed to the existence of gender dimensions of these effects. In particular, the revision of the questionnaire, aimed at better capturing unpaid work in a household business or farm, increased the female employment rate. In developing countries, the gender effects may be even more dramatic than in developed countries. Many studies have expressed concerns about the underreporting and undervaluing of women's work when using the most common methods of employment data collection (Anker, 1983; Dixon-Mueller and Anker, 1988; Charmes, 1998; Mata Greenwood, 2000). In developing countries, women workers tend to have a prominent role in agriculture and informal sector activities and, because of assigned cultural roles, may be considered by others and themselves as inactive even when they perform economic activities. In this context, it may be particularly difficult to capture women's work (Mata Greenwood, 2000).

In addition to key features of questionnaire design, different surveys adopt different approaches to designating the respondent to the questionnaire. Standard surveys in developing countries, like Household Budget Surveys (HBS), Income and Consumption Expenditure Surveys, and Core Welfare Indicator Questionnaires (CWIQ) typically ask the household head employment questions about all household members. However, proxy informants may not always provide accurate information and this can cause biases in estimation of employment (Hussmanns, Mehran, and Verma, 1990). An alternative approach is to ask each household member above a certain age directly as in the Living Standards Measurement Study surveys (LSMS) (Glewwe and Grosh, 2000) and in Labor Force Surveys (LFS). Requiring all individuals to self-report makes the fieldwork quite burdensome and expensive, creating a trade-off between the accuracy of the information and the cost to obtain it.

Most survey experiments[6] that study the effects of using proxy informants in lieu of self-respondents on employment statistics are from developed countries. In a study for the U.K., Martin and Butcher (1982), in comparing the answers of husband and wife, found that employment variables had less than a 10 percent disagreement rate, while approximately 20 percent of the proxies did not know the income of their spouse. In a similar U.S. survey, Boehm (1989) found that self and proxy responses resulted in the same labor force classification 83 percent of the time. However, this study was based on a small

---

6. Experimental studies are especially useful in assessing the "true effect" of using proxy vs. self-respondents. Non-experimental studies tend to suffer from the problem of self-selection (Hill, 1987; Moore, 1988) – that is, proxies may be individuals who happen to be at home. These proxy informants will typically have different characteristics than those who are absent from the household and those characteristics are generally correlated with the type of information that it is collected.

sample of 84 individuals from a group of participant volunteers. In general, the little experimental evidence and the non-experimental studies indicate that self-respondents produce higher household and person non-interview rates, but proxies produce higher item non-response rates, especially for wages and income variables (Biggs, 1992). The use of proxies may amplify recall errors or affect the reporting of hours of work, especially in the case of irregular or multiple activities (Hussmanns, Mehran, and Verma, 1990). Moreover, the use of proxies is also considered to be a potential source of gender bias in a context where women's participation in economic activity may be underestimated (ILO, 1982). In their study of proxy reports in the United Kingdom, The Office for National Statistics (2003) found that no one proxy informant is best placed to provide reliable proxy information for all questions. Moreover, they reject the notion of an "ideal" proxy informant in terms of personal characteristics given the variation across households. The reasons why there could be discrepancies between proxy and self-reports are reviewed in Blair et al. (1991). Experiments they conducted to analyze the strategies used by individuals to self-report or proxy report a specific event, opinion, or behavior indicate that characteristics of the questionnaire as well as individual characteristics of the self- and proxy-respondent affect the strategies used to respond and the convergence of their answers. Unfortunately their experiments do not relate to employment issues.

III. The Survey Experiment

The survey experiment conducted and analyzed here seeks to inform the method by which labor statistics are collected in household surveys in low-income countries, and, therefore, the information base for analytical work on employment. Employment is defined as time spent in an economic activity, regardless of a wage associated with it or its formal or informal nature. In this study, working includes time spent in any work for pay (as wage or salaried worker), profit (as employer, self-employed, or own-account worker), or family gain (as paid or unpaid worker in a family farm or family business). It does not include domestic work such as housekeeping, child rearing, and preparing meals – which are not comprised within the System of National Accounts (SNA) production boundary. Because of the reasons indicated in Section 2, unemployment is not a labor market measure here, as it would require a specific conceptual and methodological approach.

The survey experiment was designed and implemented to focus on two key dimensions of labor survey design: the level of detail of the questionnaire (specifically the use of screening questions to establish employment status) and the type of informant.

To investigate the impact of screening questions, a detailed and a short labor module was developed. The short labor module reflects the approach in shorter questionnaires, such as the Core Welfare Indicator Questionnaire (CWIQ).

Many countries regularly field CWIQ-type surveys (such as Welfare Monitoring Surveys), especially with increasing demands to produce sub-regional household survey statistics. This shorter module is often used to generate statistics with a higher frequency, for example with annual regularity, in lieu of complex multi-topic household surveys. The detailed labor module reflects the approach in longer questionnaires typically used in multipurpose household surveys, such as the LSMS.

In this survey experiment, the detailed module differed from the short module in two ways: in the set of screening questions to determine employment status and in asking about second and third jobs. Here the focus is on the effect of including screening questions. The detailed module starts with three questions to determine employment status: specifically, whether the person has worked for someone outside the household (as an employee), whether s/he has worked on the household farm, and whether s/he has worked in a non-farm household enterprise. For each of these three questions, the response is "yes" or "no." These questions were asked with respect to the last 7 days (the reference period for identifying those who are "employed") and, if the person has not reported to work in the last 7 days, the questions are asked with respect to the last 12 months. In the short module, there was only one question to determine the employment status with respect to the last 7 days: whether s/he did any type of work, with as response also "yes" or "no." As in the detailed module, the question was asked twice – for the last 7 days and the last 12 months. Annex Table 1 presents the questions to determine whether the individual is employed. The complete short and detailed employment modules are reported in Bardasi et al. (2010).

In the second dimension of the experiment, there is variation as to whether questions were asked directly to the subject or to a proxy informant. Response by proxy rather than individuals themselves reflects the common practice to interview an informed household member (often the household head or spouse), rather than each individual him or herself. In practice proxy informants are often used when individuals are away from the household or otherwise unavailable in the time allotted in an enumeration area to conduct interviews. In the survey experiment, the proxy informant was randomly chosen among household members at least 16 years old.[7] This age threshold reflects common practice in fieldwork to choose an adult to be a proxy informant (for children or adults) in the household. The proxy informant is thus either the head of household, spouse of the head, or an older child or relative living in the household. The persons selected to be the proxy informants then

---

7. The Tanzanian CWIQ 2006 data indicate that the average Tanzanian household has between two and three adults who could serve as a proxy with a minimum age of 16. This informed the design of our survey, and, in fact, our sample households had 2.5 members 16 years and older.

reported on themselves and on up to two other randomly selected household members age 10 or older.[8] In this paper, the responses, either proxy or self-reported, of those who are age 16 and above (which are defined as 'adults') are analyzed. In actual implementation of surveys, proxy informants are not randomly chosen, but are normally selected by interviewers on the basis of their knowledge and availability. In this sense, the experiment did not exactly mimic the actual conditions that result in proxy responses in household surveys. However, by randomly selecting proxy informants and using the information about the relationship between the proxy informant and the subject, this study can assess whether different types of proxy informants give different types of responses.[9] However, because a typical survey does not generally identify the proxy in relation to the person for whom the information is collected, the study cannot determine what the results imply in terms of potential "bias" of a typical survey due to the use of proxies.[10]

The assumption is that the self-reported information is more accurate than proxy reports. However, it is not tested whether this is true and specific reasons for the potential discrepancy cannot be identified. For example, if proxy informants report lower participation in employment, one cannot differentiate between explanations such as (1) proxies are not fully knowledgeable of the employment activity of the other household members, either because

8. Random selection of proxies was conducted in the field by the enumerator who first listed, in each household, all eligible proxies (all household members aged 16 or older) in a Table (let's call it Table A) on a proxy selection questionnaire page in the same order they appeared in the household roster. Table A was then matched with Table B, generated uniquely for each questionnaire, listing in the first column a sequence of numbers from 1 to N, where N was the total number of eligible proxy respondents in that household, and in the second column a randomly generated number in the range (1, N). The proxy respondent was chosen by selecting the individual ordered Mth in Table A, where M was the random number associated with the Nth row in Table B. The selection of the household members to be responded for by the proxy was made using a similar procedure, after excluding the selected proxy from the list of eligible members (aged 10 and older). The random selection of respondents in the self-reported sample was also made using the same procedure, but simplified to only one step.

9. There are two other reasons why the survey experiment was designed to select proxies at random. First, this design attempts to remove the influence of interviewer effects, since better interviewers will select better/more appropriate informants (and, in our view, our interviewers were well above average and had greater supervision than in a typical survey). Second, the structure of the field work suggested that if not randomly assigned, the data would be better from proxy informants than a "normal" survey because the teams were in the enumeration area for 17 days to conduct a simultaneous consumption survey experiment allowing for more time to locate the best informant.

10. An alternative research design to assess the effect of proxies would have been to interview two members of the household who report on their own labor activities and proxy report on the other. This design was not implemented because it proved to be too difficult to ensure proper implementation for a medium to large sample. After consultation with counterparts in Tanzania, it was concluded that it would be difficult to assure that proxy and self-responses would be independent and would remain unaffected by the knowledge that another household member reports on the same information, given the normally social nature of an interview. The specific concern was that the design (and open communication about this design within the village) would trigger either a coordinated response by household pairs and/or accommodation of response to the other's expectations, which would introduce potentially much larger (unobserved) respondent biases.

individuals hide their employment participation from other members, or simply because it is difficult to "keep track" of what others are doing, especially in large households; (2) proxies tend to have a "low opinion" of other household members and are likely to think that what they do does not qualify as work even when it does; or (3) the opposite, proxy informants are more likely to respond objectively and it is the individual who overstates his or her employment to make it appear that s/he works because "it looks better."

Although proxy informants and self-reporting are both commonly used, the detailed self-report questionnaire is generally considered to be the "best practice" approach of household surveys. The use of multiple questions to determine whether the subject is employed or not is recommended by the ILO, especially when some categories of workers (especially casual workers, unpaid family workers, apprentices, women engaged in non-market production, workers remunerated in-kind, etc.) may not be able to correctly interpret a question about "any type of work" as referring to their situation (Hussmanns, Mehran, and Verma 1990). The focus of this analysis is therefore whether short questionnaires provide the same information as detailed ones, and whether responses by proxy informants deviate from self-responses.

For those identified as working in the last 7 days, either through the set of three questions (in the detailed module) or through the single question (in the short module), information on the occupation, sector, employer, hours, and wage payments was collected for the main job. These questions are identical across assignments. Participation in domestic duties, while conceptually not included in the definition of employment, is commonly collected in surveys. This is usually done by adding domestic duties as a possible answer to the question about the main sector of activity and this approach was followed in both the short and detailed modules.

For all the survey assignments, in addition to the labor module, the questionnaire also included six other modules: household roster, assets, dwelling characteristics, land, food consumption, and non-food expenditures. In the detailed and short questionnaire, the questions followed the same sequence; identical types of questions follow the same phrasing and recall periods are the same.

From an analytical perspective, the objective is to assess the effects of the change in survey assignment (presence of screening questions and type of respondent). The design of our experiment introduced an imbalance in the composition of the proxy and self-report experimental groups with respect to several demographic characteristics. Proxy informants can exist only in households with at least one individual aged 16 or older and at least another one aged 10 or older; moreover, the random procedure to select proxy informants, individuals to be reported for by proxy, and self-reports is such that similar individuals have different probabilities of being selected in the two samples. We addressed the former problem by retaining for our analysis only those

households with at least two persons eligible to be a proxy informant (two persons 16 and older). The second problem was addressed by using survey weights calculated as the inverse of the selection probability. If M is the number of household members aged $10+$, the probability of being selected to self-report (in the self-report households) is 2/M if $M > 1$ and 1 if $M = 1$. In proxy households, the probability of being selected as a proxy informant (and thus also be a self-report) is $p = 1/L$, where L is the number of household members aged $16+$. The probability of being selected as an individual responded for by a proxy informant corresponds to the probability of not being selected as a proxy times the probability of being selected out of all remaining individuals eligible to be responded for, that is $w = (1\text{-}p) \times r$, where $r = \min[1, 2/(M\text{-}1)]$. After appropriately defining the samples to be compared and weighting each observation for the inverse of the probability of being selected, means can be compared across samples.

Because questions on hours, earnings, and sector are identical across assignments, variations in statistics across survey assignments are not due to question wording. However, the response to labor force participation determines whether statistics on those other dimensions are collected at all for the individual (in other words, these statistics are conditional on the individual being classified as employed). In the case of self-respondents, the screening questions that differentiate the start of the short and detailed modules entirely explain variations in selection into employment and therefore variations in hours, earnings, and sector statistics. In the case of proxy informants, variations in statistics for these other outcomes derive from both the quality of reporting by the proxy informant on a specific variable (e.g., how well the wife knows how many hours her husband works) and the accuracy of reporting on employment status (if the husband does not report that his wife works, then he will not be asked her hours). Only the latter is a selection issue.

IV. DATA AND CONTEXT

The survey experiment was implemented in Tanzania, which has different types of labor market surveys, including CWIQs, LFSs, and multi-purpose household surveys, like the Household Budget Survey (HBS). The survey experiment conducted was the Survey of Household Welfare and Labour in Tanzania (SHWALITA). The field work was conducted from September 2007 to August 2008 in villages and urban areas from 7 districts across Tanzania: one district in the regions of Dodoma, Pwani, Dar es Salaam, Manyara, and Shinyanga and two districts in the Kagera region. The sampling is a two-stage design in each region. First, villages (or urban clusters) were randomly selected proportional to their population size. Second, 12 households were randomly

selected from a household listing in each sample village (urban cluster).[11] Three of the selected 12 households were then randomly assigned to each of the four survey designs. The total sample is 1,344 households (with two of these households being replacement households selected from the original listing exercise for two households that refused to participate), with 336 households randomly assigned to each of the four survey assignments. Although the sample of 1,344 is not designed to be nationally representative of Tanzania, the districts were selected to capture variations between urban and rural areas as well as along other socio-economic dimensions.

The basic characteristics of the sampled households generally match the nationally representative data from the Household Budget Survey (2006/07) (results not presented here). Household interviews were conducted over a 12-month period but, because of small samples, the survey assignment effects across seasons (such as harvest time with a peak in labor demand and dry seasons with low demand) are not explored. The random assignment of households is validated when examining a set of household characteristics (results not presented here, but available in Bardasi et al., 2010).

The individuals are classified on the basis of the survey assignment that they actually received. An individual's actual survey assignment is the result of the initial assignment of their household among one of the four survey assignments, whether the individual is selected to be a proxy informant or a self-report, and whether the self-report or proxy assignment is realized. In the case of the self-report modules, up to two persons age 10 or older are randomly selected to self-report. If a person randomly selected to self-report are unavailable, an alternative person is selected at random. In the case of proxy assignment, one person in the household age 16 or older is selected to self-report (to maximize the number of observations in the sample) and to proxy report on up to two random household members. Because the survey experiment highly emphasized the importance of avoiding proxies, the project was fairly successful at completing self-reports when assigned. In about 5 percent of the cases, the team was unable to interview a person selected for self-report and used a proxy informant instead. The results presented in this paper are unchanged if the observations which deviated slightly from the planned design are excluded. In this paper, the focus is on the sample of subjects age 16 and older; issues related to child labor (age 10-15) are examined in another study (Dillon et al., 2010). We further restrict the sample to households with at least two persons

---

11. The selection of a fixed number of 12 households for each village does not reflect the different size of the villages. This issue should not be a concern because (1) the sample is not meant to be representative of either the whole country or meaningful parts of it, and (2) the focus of the paper is a comparison across 'similar' groups of individuals with the purpose of highlighting differences in statistics rather than discussing the levels and meaning of those statistics for the Tanzanian labor market. For this reason, it was decided not to correct the household weights to reflect the unequal household selection probabilities across villages given that this correction would be irrelevant to the analysis.

TABLE 1. Individual and household characteristics, by survey assignment

	Individual survey assignment				
	Detailed	Detailed	Short	Short	F-test of equality of coefficients across groups
	Self-report	Proxy	Self-report	Proxy	
Individual characteristics					
Female (%)	54.8	50.3	50.0	50.4	0.309
Age	35.5	36.0	36.2	37.0	0.657
Highest school grade attended	4.9	4.8	4.9	5.0	0.922
Married (%)	63.1	65.7	68.0	61.6	0.161
Household characteristics					
Head: female (%)	14.6	15.2	14.3	16.2	0.781
Head: age	47.9	48.1	48.8	49.6	0.457
Head: highest school grade attended	4.8	4.8	4.8	4.8	0.993
Head: married (%)	83.9	82.8	83.9	81.8	0.674
Household size	6.6	6.5	6.6	6.9	0.438
Adult equivalence household size	5.4	5.3	5.3	5.5	0.431
Share of members less 6 years	18.3	17.8	18.2	17.8	0.798
Share of members 6-15 years	25.1	23.9	24.3	23.9	0.443
Number of adults 16+ years	3.5	3.6	3.5	3.7	0.293
Concrete/tile flooring (non-earth) (%)	25.8	27.7	23.6	24.5	0.669
Main source for lighting is electricity/ generator/solar panels (%)	11.2	11.0	11.6	13.5	0.723
Owns a mobile telephone (%)	33.4	35.3	33.7	37.4	0.384
Bicycle (%)	50.1	49.5	54.4	54.5	0.615
Owns any land (%)	82.3	82.4	84.4	84.3	0.859
Acres of land owned (including 0s)	3.7	3.4	3.7	3.7	0.526
Urban (%)	33.5	33.2	31.9	33.0	0.925
Month of interview (1 = Jan, 12 = Dec)	5.6	5.4	5.5	5.5	0.673
N of individuals	734	373	720	366	

Notes: See NBS (2002) for details on the adult equivalence scales. The F-test tests the equality of coefficients across the groups by regressing the group indicators on the household characteristics with clustered household standard errors. Includes person-weights defined in the text.

eligible to be a proxy informant (two persons 16 and older). Summary statistics for the sample are presented in Table 1.

V. RESULTS

The presentation of the results of the experiment is divided into two parts. In the first part, differences across the survey assignments are examined for key employment statistics on the individual's main activity: labor force participation, weekly hours, daily earnings, the sector of work, and type of work (employment status). The statistics (averages) both between the short module and the detailed module, and between responses given by proxy and self-reported responses are compared. Because a slight unbalance across experimental groups persists even after weighting for unequal selection probabilities

(probably due to the relatively small sample sizes of the groups – see Table 1), we decided to run regressions (with weights) to fully control for discrepancies in the composition of the experimental groups introduced by the survey design:

$$y_i = \alpha + \beta_S S_h + \beta_P P_i + \lambda X_i + \gamma D_h + \varepsilon_h \qquad \text{(Eq.1)}$$

where y_i are the different labor statistics (like labor force participation, labor supply, earnings, and occupational choice) for the ith individual, S_h is an indicator variable for the short questionnaire treatment of individuals in household h, P_i is an indicator variable for the proxy treatment of individual i in household h, X_i is a vector of individual and household characteristics for the ith individual, D captures district indicators, and ε is the stochastic error term, which is randomly distributed across households. The marginal treatment effects are estimated using standard models (OLS, probit, and multinomial logit).

In the second part, the impact of the characteristics of the proxy informants on the employment statistics are examined, specifically whether there are "types of proxy informants" who generate statistics that are closer to self-reports.

Differences in Labor Statistics across Survey Assignment

Table 2 presents the findings, disaggregated by gender, for employment, weekly hours, and daily earnings. In each case, the difference in means across survey assignments is tested using a t-test. Row 1 of Table 2, for instance, reports the employment rate of men from the short module (91.2 percent) and from the detailed module (85.7 percent), and finds that the difference (5.5 percentage points) is statistically different from zero at the 1% level.

When looking at the employment rates based on the informant's classification (i.e., derived from the one question "Did you do any type of work in the last 7 days" in the short module, and from the three screening questions specifying three main groups of economic activities in the detailed module), the short module produces *higher* employment rates than the detailed module, for both men and women (Table 2, top panel). This result is in contrast with what was expected *a priori* – that a generic and vague question about "work" would miss people in marginal activities and activities with no remuneration. However, after re-classifying domestic duties into "no work" as per the ILO definition, shifts in employment rates are observed, especially for women (Table 2, second panel). For men, the decrease in the employment rate is small in the short and even smaller in the detailed questionnaire, so that there is no statistically significant difference in the eventual employment rates produced by the two survey instruments (88 and 85 percent, respectively). For women, however, there is a substantial number of reclassifications needed when using the short questionnaire. Because a very large percentage of women gets classified as "working" but is carrying out domestic duties, the percentage in

TABLE 2. Labor statistics by survey assignment and sex

	A.			B.			Number of observations
	Short	Detailed	Diff	Proxy	Self-rep	Diff	
Participation in employment (informant's classification)[a] (%)							
Men	91.2	85.7	5.5***	81.1	92.7	−11.6***	1062
Women	89.5	81.5	8.0***	84.4	85.8	−1.4	1131
Participation in employment (after reclassification of domestic duties)[b] (%)							
Men	87.9	85.0	2.9	78.5	91.1	−12.6***	1062
Women	75.0	80.1	−5.1*	76.1	78.5	−2.4	1131
Weekly hours last week unconditional on employment [b] (mean)							
Men	32.8	30.5	2.3*	28.6	33.5	−4.9***	1059
Women	24.6	25.9	−1.3	25.5	25.2	0.3	1128
Weekly hours last week among working (if employment = 1) [b] (mean)							
Men	37.4	35.9	1.5	36.4	36.8	−0.4	924
Women	32.9	32.3	0.6	33.5	32.1	1.4	880
Conditional daily earnings (Tshillings) (if employment =1 and earnings > 0)[b] (mean)							
Men	5,064	3,871	1,193	5,729	3,696	2,033**	168
Women	4,803	4,505	298	5,211	4,255	956	82

Notes: 'Diff' indicates the difference between the averages reported in the two preceding columns. Includes person-weights defined in the text. *** indicates statistical significance at 1%, ** at 5%, * at 10%.

[a] For the short questionnaire, this is the percentage of those who answer "Yes" to Question 1 (Annex Table 1, first column); for the detailed questionnaire, this is the percent of "Any yes" to Questions 1, 3, and 7 (Annex Table 1, second column).

[b] Participation in employment after re-classifying those who indicated domestic duties as their main work activity (Annex Table 1 Question 4 in the first column and Question 9 in the second column, for the short and detailed questionnaire, respectively) into non-employment, according to the ILO definition.

employment according to the short questionnaire decreases from almost 90 to 75 after correct reclassification. The variation is much smaller in the detailed questionnaire, with only a handful of women re-classified as non-working. As a result, eventual female employment based on the short module becomes about 5 percentage points *lower* than in the detailed module.

Using proxy informants generates male employment rates that are more than 10 percentage points lower than when using self-reports. After re-classifying domestic duties into no-work, the difference between the proxy and self-reported male employment statistics remains large and statistically significant (about 13 percentage points lower for proxy reports). Comparing the first and second panels of Table 2, for both proxy informants and self-reports, employment rates are "inflated" when domestic duties are not re-classified. Female employment rates, by contrast, do not differ substantially between proxy informants and self-reports, although both are lower after the reclassification of domestic duties into non-work.

When the variation in conditional and unconditional weekly hours are examined across survey experiments, the average unconditional number of

weekly hours based on reports by proxy informants is significantly lower for men (about 5 hours less per week on average), but not for women; however, for employed men proxy informants report the same weekly hours as self-reports (about 36.5 per week). This result is driven by the propensity of proxy informants to report a much lower participation in employment for men. If marginal jobs are those that are being underreported with the short or proxy surveys, one would expect to see lower hours among the employed for these two groups with respect to the means generated by the detailed questionnaire and the self-reports. This is what is observed for men, for whom conditional hours are larger in the short (37.4 per week) than in the detailed module (35.9 per week – see fourth panel of Table 2), while for women the difference is smaller (32.9 and 32.3 hours per week, respectively). Results for average hours are similar when employment of subjects working more than 40 hours per week are examined (results not presented).

There are differences in daily earnings between survey assignments, but because of the small number of observations (most individuals employed in agriculture or as unpaid family members do not derive earnings from their activity) they are mostly not statistically significant. The detailed module tends to produce lower average earnings for both men and women, and self-reporting also generates lower earnings for both men and women; however only for men the difference between proxy and self-reporting is especially large and statistically significant.

In Table 3, the distribution across main activities by assignment is presented. Activities are classified into four categories. Employed individuals are distributed between agriculture and other sectors.[12] The category "domestic duties" (included as a possible answer alongside other industrial sectors of activity) is kept as a separate category.[13] The fourth category in Table 3, "no work" corresponds therefore to the informant's definition.

Panel A in Table 3 shows that the higher male employment rate in the short module observed in Table 2, although not statistically significant, stems from men being less likely to be in "no work" (about 6 percentage points). By contrast, the lower female employment rate in the short module results primarily from a large "participation" in domestic duties as women's main activity. In the case of the detailed module, women are much more likely to be classified as "not working" than in domestic duties; in the detailed module, only 1

---

12. The non-agricultural sectors are too small to consider in a disaggregated manner. These include: mining/quarrying/manufacturing/processing, gas/water/electricity, construction, transport, buying and selling, personal services, education/health, and public administration. Buying and selling activities are the most frequently reported of these activities (4-7 percent, depending on the sub-group).

13. Although "domestic duties" is listed as a potential "sector of main activity," the interviewers received clear instructions to include any domestic duties contracted outside the household in the category "personal services" (counted as employment) and classify under "domestic duties" only domestic and household work done for the household where individuals live. Careful debriefing confirmed that these guidelines were strictly followed.

TABLE 3. Sector of main activity by survey assignment and sex

	Men			Women		
A. Short or Detailed	Short	Detailed	Diff	Short	Detailed	Diff
Main activity^						
Agriculture	65.4	59.1	6.3**	64.9	67.2	−2.3
Other sectors	22.6	25.9	−3.3	10.1	12.9	−2.8
Domestic Duties	3.3	0.7	2.6***	14.5	1.4	13.1***
No work	8.8	14.3	−5.5***	10.5	18.5	−8.0***
N	539	523		547	584	
Main activity among workers^						
Agriculture	74.3	69.5	4.8	86.6	83.9	2.7
Other sectors	25.7	30.5	−4.8	13.4	16.1	−2.7
N	476	451		413	470	
B. Proxy or Self-report	Proxy	Self-rep	Diff	Proxy	Self-rep	Diff
Main activity^						
Agriculture	53.7	67.2	−13.5***	63.6	67.5	−3.9
Other sectors	24.9	23.8	1.1	12.5	11.0	1.5
Domestic Duties	2.6	1.7	0.9	8.2	7.3	0.9
No work	18.9	7.3	11.6***	15.6	14.2	1.4
N	360	702		379	752	
Main activity among workers^						
Agriculture	68.3	73.8	−5.5*	83.6	86.0	−2.4
Other sectors	31.7	26.2	5.5*	16.4	14.0	2.4
N	286	641		292	591	

Notes: Includes person-weights defined in the text. Other sectors are specifically listed on the questionnaire and include mining/quarrying, manufacturing/processing, gas/water/electricity, construction, transport, trading, personal services, education/health, public administration, and other. 'Diff' indicates the difference between the averages reported in the two preceding columns. *** indicates statistical significance at 1%, ** at 5%, * at 10%.

^Percentages by group may not sum to 100 due to rounding.

percent of women are classified in domestic duties, about 13 percentage points less with respect to the short module. This suggests that the additional questions at the beginning of the detailed employment module succeed in filtering out the large majority of individuals who would otherwise be classified under domestic duties. These three questions – explicitly mentioning and exemplifying farm work, wage work, and work in household enterprises – frame the notion of work to exclude domestic duties in the minds of respondents. The sector composition of employment (for those employed; i.e., the distribution between agriculture and other sectors) is not significantly affected by the short or detailed modules for either men or women.

Comparing proxy and self-reports, the distribution across sectors is affected for men but not for women (Table 3 panel B). Proxy informants report lower employment for men as a result of a higher percentage of "no work" (11.6 percentage points difference between proxy and self-reporting). Interestingly, the decline in male employment when moving from self-report to proxy is almost entirely accounted for by lower participation in agriculture (13.5 percentage

TABLE 4. Employment status among employed by survey assignment and sex

	Men			Women		
A. Short or Detailed^	Short	Detailed	Diff	Short	Detailed	Diff
Paid employee	13.0	20.7	− 7.7***	5.0	12.2	− 7.2***
Self-employed, with employees	3.4	4.9	− 1.5	1.1	0.9	0.2
Self-employed, no employees	61.0	57.7	3.3	23.8	21.8	2.0
Unpaid family worker	22.5	16.7	5.8**	70.1	65.1	5.0*
N	474	447		411	467	
B. Proxy or Self-report	Proxy	Self-rep	Diff	Proxy	Self-rep	Diff
Paid employee	15.9	17.1	− 1.2	9.0	8.9	0.1
Self-employed, with employees	2.8	4.8	− 2.0	0.9	1.0	− 0.1
Self-employed, no employees	62.7	57.8	4.9	24.4	21.9	2.5
Unpaid family worker	18.6	20.3	− 1.7	65.7	68.2	− 2.5
N	285	636		291	587	

Notes: 'Diff' indicates the difference between the averages reported in the two preceding columns. Includes person-weights defined in the text. *** indicates statistical significance at 1%, ** at 5%, * at 10%.

^Percentages by group may not sum to 100 due to rounding

The sample excludes respondents who indicated domestic duties as their main work activity (see Table 2 note b).

points less as reported by proxy), while participation in other sectors is almost unaffected. This suggests that a large proportion of men who classify themselves as working in agriculture are not recognized as working by other adult household members. For women, significant differences between proxy and self-reports are not observed.

The distribution of those in employment by status can also be examined with the data: paid employees, self-employed with employees, self-employed with no employees (often farmers), and unpaid family worker. The results are presented in Table 4. Comparing the short and detailed modules, significant variations in several categories are observed, with the short module producing less paid employment for both men and women (more than 7 percentage points difference), compensated by more self-employment without employees (although not statistically significant) and more unpaid work (statistically significant) for both men and women. The difference in paid employment is large and unexpected – one would have assumed that equating paid employment with work should have been easy even when the concept of work is not precisely defined. The impact of asking by proxy is not very large and never statistically significant, and results in higher self-employment without employees and lower percentages of unpaid family workers for both men and women. In this case we do not observe substantial variations in the proportion of workers who are paid employees.

We estimate equation 1 to measure the marginal effect of each treatment, including the weights described above and controlling for observable characteristics. Controls include individual characteristics of the subjects (age, sex, and

TABLE 5. Probit and OLS regressions of labor statistics by survey assignment and sex

	Participation in employment		Conditional weekly hours		Conditional daily earnings	
	(1) Men	(2) Women	(3) Men	(4) Women	(5) Men	(6) Women
Short	0.015	−0.044*	0.122**	0.055	0.142	0.081
	(0.021)	(0.027)	(0.052)	(0.050)	(0.144)	(0.311)
Proxy	−0.112***	−0.012	−0.066	0.042	0.183	0.256
	(0.024)	(0.028)	(0.056)	(0.050)	(0.161)	(0.246)
N	1,062	1,131	924	880	168	82

Notes: Standard errors in parentheses. Includes person-weights defined in the text.

Columns 1-2: marginal effects of a probit regression for labor force participation are reported. See Table 2 note (b) regarding excluding domestic duties as employment.

Columns 3-4: coefficients of an OLS regression on log hours are reported.

Columns 5-6: coefficients of an OLS regression on log daily earnings are reported.

Other, covariates included but not presented are characteristics of the subject (age, gender, education) his/her household (household composition and assets) and district fixed effects.

*** indicates statistical significance at 1%, ** at 5%, * at 10%.

education), household characteristics, and district dummies. The results reported in Table 5 show the large effect of proxy informants on the male employment rate – 11 percentage points less by proxy than self-reporting, the omitted category. By contrast, in the case of women, it is the short questionnaire that is responsible for 4 percentage points less employment with respect to the detailed module (employment in this table correctly excludes domestic duties). These results are similar to the findings presented in Table 2 (second panel). The results by urban and rural location are also examined (results not presented here).[14] A lower employment rate for men using proxy informants is found in rural areas, but not in urban areas, where employment rates for men do not differ between proxy and self-reports. By contrast, the lower employment rate for women obtained with the short module is driven by the urban areas, while in rural areas no difference between the short and detailed modules was found.

Columns 3 and 4 in Table 5 report the estimates of an OLS regression for (log) weekly hours. The short questionnaire is associated with higher average hours for men (about 12 percent more), which may be explained by some underreporting of "short" or marginal jobs in the short questionnaire. The use of proxy informants, by contrast, does not affect average weekly hours.[15] To assess whether this is the correct interpretation, probit models are estimated for the probability of working at least a certain number of hours (20, 30, and 40

14. In urban locations in our sample, agriculture is still a major sector, accounting for about one-third of all jobs.

15. By urban and rural location, these results are only found among rural households.

weekly hour thresholds), conditional on working. Including the same covariates and treatment effects as in Table 5 in the regressions, the coefficient of the short module is significant for the 20-hours threshold for both men and women; specifically, using the short rather than the detailed module produced a 7 percentage point higher probability of jobs longer than 20 hours for those working. However, the coefficient is smaller and not significant when adopting higher thresholds (with the exception of men for the 30 hour threshold which is significant at 10 percentage points). This result can be interpreted as a higher propensity of the short module to "miss" shorter jobs, while longer jobs (with the number of hours equal to or larger than 30 hours/week, for example) are equally likely to be reported by both the short and detailed modules (and this is true for both men and women). Finally, with respect to daily earnings, there are no significant differences (see Columns 5 and 6 of Table 5).

Using a multinomial logit, the effects of the survey assignment on the distribution across three employment categories is estimated: agriculture, other sectors, and the omitted category "not in employment", including also domestic duties. The results shown in Table 6 confirm the large effect of proxy informants on male participation in agriculture (17 percentage points less than self-reports) with respect to "not in employment". The different survey instruments do not generate significant differences for women.[16]

Following a similar estimation approach for employment status (see Table 7), controlling for proxy assignment, the short module decreases the probability of men and women of being in wage employment compared with unpaid work, and increases the probability of men being self-employed without employees (albeit the coefficient is not statistically significant). Relying on proxy informants does not impact status in the main job for either men or women.

The Impact of the Characteristics of Proxy Informants

Because proxy informants were randomly selected, the dataset contains some variability of "proxy types", which can be used to address the question of whether the proxy's characteristics affect the proxy's responses to the employment questions.[17] Ideally, these data would be used to identify proxy informants' characteristics that produce the "best" results, in order to advise survey methodologists on how to choose the proxy informant. In the absence of better

16. As a robustness check, the equations presented in Tables 4 and 5 are re-estimated dropping from the sample the responses of proxy informants who self-reported on themselves before reporting on other household members. The results are virtually unchanged – the only difference is that the coefficient of the short questionnaire in the conditional hours equation is no longer statistically significant for either men or women. This effect may be partly due to the reduction in statistical power because of the small sample size. These equations were also re-estimated excluding the set of controls, and the results are unchanged as expected given the randomized design of the survey assignment.

17. Analysis of the impacts on hours, earnings, and status in employment is not possible because of sample size considerations.

TABLE 6. Multinomial logit of main activity by survey assignment and sex

	Men		Women	
	Agriculture	Other Sectors	Agriculture	Other Sectors
Short	0.031	−0.010	−0.034	−0.014
	(0.037)	(0.031)	(0.039)	(0.011)
Proxy	−0.173***	0.046	−0.020	0.001
	(0.040)	(0.032)	(0.043)	(0.012)
N	1,062		1,131	

Notes: Marginal effects are reported. Includes person-weights defined in the text. Other covariates included but not presented are characteristics of the subject (age, sex, education), his/her household (household composition and assets) and district fixed effects. The three categories of the multinomial logit model are agriculture, other sectors, and "out of work" (domestic work and no work – omitted category). See note to Table 3 for an explanation of other sectors. *** indicates statistical significance at 1%, ** at 5%, * at 10%.

TABLE 7. Multinomial logit of employment status in main job in the last 7 days by survey assignment and sex

	Men			Women		
	Wage employee	Self-employed with employees	Self-employed, no employees	Wage employee	Self-employed with employees	Self-employed, no employees
Short	−0.066**	−0.001	0.043	−0.062***	0.000	0.009
	(0.031)	(0.001)	(0.036)	(0.023)	(0.000)	(0.027)
Proxy	−0.037	−0.001	0.053	0.009	0.000	0.005
	(0.027)	(0.001)	(0.034)	(0.022)	(0.000)	(0.028)
N	921			878		

Notes: Relative risk ratios are reported. Includes person-weights defined in the text. Other covariates included but not presented are characteristics of the subject (age, gender, education), his/her household (household composition and assets) and district fixed effects. The multinomial logit model is used where the four categories include wage employee, self-employed with employees, self-employed without employees, and unpaid family worker (the omitted category). *** indicates statistical significance at 1%, ** at 5%, * at 10%.

criteria, the "best" is defined as the proxy report that converges to the self-report, i.e., a concordant answer (this is the definition adopted by other researchers, including for example Blair et al., 1991). Unfortunately, both a proxy response and a self-report for the same individual is not available – in which case, one could analyze the characteristics of the proxy informants whose answers approach the self-report for the same person. Another complication arises from the fact that the characteristics of the proxy informants are not unrelated to the characteristics of the individuals on whom they report. For example, a test of whether parents are "better" proxies than other adult household members can only be carried out for young people who reside with at

least one parent and at least one other adult other than a parent. In comparing the reports of proxy informants with self-reports, therefore, the sample should be selected in such a way to create the correct comparison group, to make sure that the differences in employment statistics only reflect the "proxy effect" and not genuine differences in employment behavior driven by differences in household composition.

The first strategy to address this concern is to control for observed individual and household characteristics – household composition variables being especially important. However, unobserved traits related to sample selection may still be a concern. The second strategy, again, controls for observed characteristics, but also conditions the sample to the "relevant" one. This is the sub-sample of those households whose composition allows for all possible selections of the informant considered in each specific regression (the econometric motivation is one of a non-parametric matching between the proxy and self-report households restricting the matched sample to the common support). For example, when examining the impact of the gender of the proxy informant, only select those households that have at least one adult male and one adult female in addition to the subject (an adult of each gender who could possibly be selected to provide proxy information). Unfortunately, this strategy results in small sample sizes and a loss of precision of the estimates. Instead, the focus is on four specific characteristics of proxy informants: gender, education, marital status, and headship.

The results are reported in Table 8. Note that the sample changes with each regression. In all regressions, the small and not-significant effect of proxy informants on female employment rates is confirmed; the substantial and statistically significant effect of proxy informants on male employment rates is also confirmed. The size of the proxy effect is now generally larger, but these coefficients are not directly comparable with those presented in the previous regressions, because they refer to a different subsample of individuals in each column.[18] All proxy informants substantially underestimate male employment participation, but proxies who are better educated and (female) spouses do better than others – the bias in both cases is "only" about 12 percentage points (for educated proxies, column 3) and 7 percentage points (for female spouses, column 5). Choosing the head of the household to report on male participation (column 7) does not produce "better" statistics than when other adults are selected as proxies. However, the extremely large coefficients estimated in this case for a non-head proxy respondent (about 23 percentage points) is very likely driven by the characteristics of the subject. Men who are not the head and reside in a household with at least one other adult who is not the head are likely to be adult children. They may be disproportionately found in marginal, casual, or irregular employment. In both cases, the proxy may

---

18. This also explains why the impact of the short vs. detailed questionnaire, which is significant for women in the general specification, now differs across household types.

TABLE 8. Participation in employment by proxy characteristics

Subject sample:	(1) Men	(2) Women	(3) Men	(4) Women	(5) Men	(6) Women	(7) Men	(8) Women
Short	0.023	−0.051	0.025	−0.099	0.017	−0.045*	0.032	−0.049*
	(0.033)	(0.048)	(0.026)	(0.075)	(0.021)	(0.027)	(0.042)	(0.029)
Proxy	−0.166***	−0.016	−0.183***	−0.083	−0.167***	−0.029	−0.229***	−0.057
	(0.044)	(0.058)	(0.053)	(0.094)	(0.030)	(0.036)	(0.053)	(0.049)
Proxy is man	0.009	−0.018						
	(0.045)	(0.078)						
Proxy has schooling			0.062**	0.050				
			(0.025)	(0.094)				
Proxy is spouse					0.096***	0.041		
					(0.019)	(0.045)		
Proxy is head							0.015	0.063
							(0.060)	(0.046)
N	382	421	211	182	1,062	1,131	316	1,014

Notes: Standard errors in parentheses. Includes person-weights defined in the text.

Marginal effects of a probit regression for participation in employment are reported. Domestic duties are classified as non-employment (see Table 2 note (b)).

Other covariates included but not presented are characteristics of the subject (age, sex, education), his/her household (household composition and assets), and district fixed effects.

Columns 1 and 2 are restricted to subjects who reside in a household with at least 1 other adult male and 1 other adult female. Columns 3 and 4 are restricted to subjects who reside in a household with at least 1 other adult with schooling and 1 other adult with no schooling. Columns 5 and 6 are restricted to subjects who reside in a household with at least 1 spouse and one other adult. Columns 7 and 8 are restricted to subjects who are not the head and reside in a household with at least 1 other adult who is not the head.

*** indicates statistical significance at 1%, ** at 5%, * at 10%.

have poor knowledge of the male subject's participation in employment. Still, in the case of women, choosing the head does not offer any obvious advantage – heads and non-heads both do pretty well.

Cost Implications of Survey Design

Alternative survey designs will have cost implications that have to be weighed against a subjective valuation of "better" data. Using the detailed module added only a few minutes to the average duration of the interview compared with the shorter module, according to field work reports from enumerators and supervisors. The cost implication of using a detailed rather than a short module, therefore, is small. The additional cost of printing slightly longer questionnaires and the extra data entry requirement are also only marginal.

By contrast, using proxy instead of self-reports involves substantial savings. The use of self-reports increases the length of field work because more days are spent in each sample village to locate and interview respondents. Since this survey experiment was carried out in conjunction with a larger consumption expenditure experiment, it required survey teams to spend a full two weeks in each village. This ensured that self-reports were completed but, as a consequence, it is difficult to report on precisely how many extra days are needed to complete self-reports compared with proxy labor modules. However, based on field experience, a rough calculation can be made that for two days spent in a village using proxy informants, the survey team would need at least one more day to track down self-reports. This corresponds to a 50 percent increase in the length of time spent on actual field work. All variable costs of field staff (per diems, lodging costs), often the largest category of survey costs, are assumed to increase by 50 percent. Transport costs may also increase if field teams use a team vehicle to track down respondents for self-reports.

Our results make the valuation of the trade-off between better data and lower survey costs complicated. For women, proxy reporting has no effect on female labor statistics in the sample. Given the higher costs associated with self-reporting, there seems to be no substantial benefit for insisting on self-reporting for women. This is similar to other findings on survey design and child labor (Dillon et al., 2010). Men's labor statistics, however, particularly employment rates and participation in agriculture, are substantially affected by proxy reporting. Given the heterogenous effects of survey design within this sample and the magnitude of the proxy effect on men's labor force participation and probability of working in agriculture, serious reflection by survey designers with respect to the objectives and precision necessary in their surveys is paramount. For national surveys on labor activities, self-reporting appears to be the best practice. However, further research on the reliability of self-reporting itself is necessary to confirm this conjecture, as well as validation of the results in other contexts.

VI. CONCLUSIONS

Despite the importance of household survey instruments as a source for labor statistics, there is a dearth of evidence on the best practices for collecting these statistics in developing countries. The differences in survey design for national labor statistics over time within a country and across countries have serious implications for both measuring labor market outcomes and carrying out research on labor activities. This paper presents a survey experiment focusing on two key aspects of survey design: the questions to identify participation in employment and the choice of the respondent. The statistics generated from short and detailed questionnaires, and by a proxy informant and self-reporting are compared.

The findings suggest that both types of survey design decisions have statistically significant effects on labor statistics. These effects are largest on employment participation rates, but also exist for weekly hours of work, main activity, and type of work. The effects vary depending on the labor market outcome and remain significant even after controlling for individual, household, and village characteristics and sample composition. The short questionnaire captures a large percentage of women "employed" who are actually engaged in domestic duties; but reclassification of these cases into "no-work" produces a lower female employment rate than the detailed questionnaire. Moreover, it also tends to generate higher average weekly hours conditional on working, especially for men, as well as a lower share of paid employees among the employed. Response by proxy leads to substantially lower male employment rates due to lower reporting of participation in agriculture – while for women there is no effect. The discrepancies between proxy and self-reporting are reduced when the spouse is selected as the proxy and when the proxy is an individual with some education.

In this paper, only two dimensions of survey design are the focus; future work is needed to look at other issues, such as the wording of the questions. But even the (limited) results presented here provide some clear advice for survey design. First, the impacts are not consistently associated with one specific survey feature but differ for different types of individuals. Using a short rather than a detailed module produces lower employment for women but not for men; using proxies rather than asking the individual directly strongly impacts (negatively) employment for men but not for women. This indicates that the "best" approach – if it exists – may differ depending on the purpose of the survey, which determines the type of variables and the type of sample for which information is collected.

Second, the inclusion of a "domestic duties" category as a possible answer to the "main sector of activity" can produce ambiguous results, especially in a short employment module, and in particular for women. Individuals may classify themselves (or be classified) as "employed" when responding to a very direct yes/no question about working, but then *de facto* rule themselves out of

employment (based on the SNA definition of economic activity) by indicating that their main activity was being engaged in domestic duties. When using a short module, ambiguity remains as to how these individuals interpret the meaning of "employment" and "domestic duties" and whether they tend to give priority to "domestic duties" as a main activity over other activities in which they may have worked as well. Without information about other jobs, this ambiguity is left unresolved.

Third, in a low-income setting, the distinction between agricultural work and "no work" can be subject to interpretation. The proxy informants are much more likely to categorize men out of work rather than working in agriculture with respect to self-reports. Future work may want to investigate this issue further.

Fourth, given the – sometimes unexpected – variations that distinct survey designs can generate, these results underline the importance of staying with the same design if the aim is to make comparison, both over time or across contexts.

If the detailed self-reported module is best practice – as recommended by ILO guidelines – these results provide an assessment of the implications for data quality of using less expensive approaches, such as response by proxy or using a shorter module. Note that the trade-off between higher quality of data and lower implementation costs is especially driven by the choice of using proxy informants vs. self-reports, and not so much by the small number of additional questions in the detailed questionnaire. These findings suggest that there is no substantial benefit of self-reporting for women (a similar result was found for children, as discussed in Dillon et al., 2010), but there is for men, whose employment rates and distribution across sector of activity are affected by proxy reporting. The results also imply that the detailed questionnaire should be maintained as best practice approach, because using the short questionnaire does not involve relevant cost savings but may affect some of the statistics (in this case, female employment rates, male working hours, and the employment status of both men and women).

This research also points to a number of fruitful avenues for future work. A common frustration with this type of work is that there is no clear evidence of a gold standard (yet). While self-reporting is generally accepted as best practice, further research on the reliability of self-reporting itself is necessary to confirm this conjecture. The experiment analyzed the convergence of proxy and self-reports, but did not address the issue of whether any of these reports actually correspond to the "truth". Moreover, the large percentage of informants that identify "domestic duties" with "employment" or are unable to distinguish between working in agriculture and being out of work suggests that the notion of "employment" should not be taken as given, but needs to be defined and clarified, especially in low-income countries. This is as much an analytical as a methodological question. It leaves open the question of how questionnaires that have been mostly developed with reference to advanced economies can be

further improved to make them fully adequate to capture "employment" in completely different settings.

Annex Table 1. Key employment questions in the short and detailed questionnaires

Short questionnaire	Detailed questionnaire
	1. During the past 7 days, has [NAME] worked for someone who is not a member of your household, for example, an enterprise, company, the government or any other individual? YES . . . 1 (»3) NO . . . 2 (question repeated for the past 12 months – question 2)
	3. During the past 7 days, has [NAME] worked on a farm owned, borrowed or rented by a member of your household, whether in cultivating crops or in other farm maintenance tasks, or has [NAME] cared for livestock belonging to a member of your household? YES . . . 1 (»5) NO . . . 2 (question repeated for the past 12 months – question 4)
	5. During the past 7 days, has [NAME] worked on his/her own account or in a business enterprise belonging to him/her or someone in your household, for example, as a trader, shop-keeper, barber, dressmaker, carpenter or taxi driver? YES . . . 1 (»7) NO . . . 2 (question repeated for the past 12 months – question 6)
1. Did [NAME] do any type of work in the last seven days? Even if for 1 hour. YES . . . 1 (»3) NO . . . 2 (question repeated for the past 12 months – question 2)	7. CHECK THE ANSWERS TO QUESTIONS 1, 3 AND 5. (WORKED IN LAST 7 DAYS) ANY YES.. 1 ALL NO . . . 2 (»37)
3. What is [NAME]'s primary occupation in [NAME]'s main job?	8. What is [NAME]'s primary occupation in [NAME]'s main job?

(Continued)

ANNEX TABLE 1. Continued

Short questionnaire	Detailed questionnaire
(MAIN OCCUPATION IN THE LAST 7 DAYS)	(MAIN OCCUPATION IN THE LAST 7 DAYS)
a. OCCUPATION	a. OCCUPATION
b. OCCUPATION CODE	b. OCCUPATION CODE
4. In what sector is this main activity?	9. In what sector is this main activity?
AGRICULTURE. 1	AGRICULTURE. 1
MINING/QUARRYING. 2	MINING/QUARRYING. 2
MANUFACTURING/ PROCESSING. 3	MANUFACTURING/ PROCESSING. 3
GAS/WATER/ELECTRICITY. 4	GAS/WATER/ELECTRICITY. 4
CONSTRUCTION. 5	CONSTRUCTION. 5
TRANSPORT. 6	TRANSPORT. 6
BUYING AND SELLING. 7	BUYING AND SELLING. 7
PERSONAL SERVICES. 8	PERSONAL SERVICES. 8
EDUCATION/HEALTH. 9	EDUCATION/HEALTH. 9
PUBLIC ADMINISTRATION. 10	PUBLIC ADMINISTRATION. 10
DOMESTIC DUTIES. 11	DOMESTIC DUTIES. 11
OTHER, SPECIFY. 12	OTHER, SPECIFY. 12

REFERENCES

Abowd, J.M., and A. Zellner 1985. "Estimating Gross Labor-Force Flows." *Journal of Business & Economic Statistics* 3(3):254–283.

Anker, R. 1983. "Female Labour Force Participation in Developing Countries: A Critique of Current Definitions and Data Collection Methods." *International Labour Review* 122(6):709–724.

Bardasi, E., K. Beegle, A. Dillon, and P. Serneels 2010. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." World Bank Policy Research Working Paper 5192.

Biemer, P., R. Groves, L. Lyberg, N. Mathiowetz, and S. Sudman 1991. *Measurement Error in Surveys*. New York: John Wiley & Sons.

Biggs, B. 1992. "Self/Proxy informant Rules and Data Quality." Research Paper, Income Research Paper Series, Ottawa: Statistics Canada.

Blair, J., G. Menon, and B. Bickart 1991. "Measurement Effects in Self vs. Proxy Responses to Survey Questions: An Information-Processing Perspective." in P P. Biemer, R. Groves, L. Lyberg, N. MathiowetzS. Sudman (eds.) *Measurement Error in Surveys*. New York: John Wiley & Sons.

Boehm, L.M. 1989. "Reliability of Proxy Response in the Current Population Survey." in *Proceedings of the Survey Research Methods Section, American Statistical Association*.

Bound, J., C. Brown, and N. Mathiowetz 2001. "Measurement Error in Survey Data." In *Handbook of Econometrics* Vol. 5. ed. J. Heckman, and E. Leamer Amsterdam: North-Holland, Elsevier Science.

Campanelli, P., J.M. Rothgeb, and E.A. Martin 1989. *The Role of Respondent Comprehension and Interviewer Knowledge in CPS Labor Force Classification*. American Statistical Association Proceedings (Survey Research Methods Section).

Charmes, J. 1998. *Women Working in the Informal Sector in Africa: New Methods and New Data*. Paris: Scientific Research Institute for Development and Co-operation.

Dillon, A., E. Bardasi, K. Beegle, and P. Serneels 2010. "What Explains Variation in Child Labor Statistics? Evidence from a Survey Experiment in Tanzania." World Bank Policy Research Working Paper 5414.

Dixon-Mueller, R., and R. Anker 1988. *Assessing Women's Economic Contributions to Development.* Training in Population, Human Resources and Development Planning Paper number 6, Geneva: International Labour Office.

Esposito, J.L., P.C. Campanelli, J. Rothgeb, and A.E. Polivka 1991. "Determining which Questions Are Best: Methodologies for Evaluating Survey Questions." *Proceedings of the American Statistical Association* (Survey Research Methods Section).

P. Glewwe, and M. Grosh (eds). 2000. *Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Development Study.* Oxford University Press (for the World Bank).

Guarcello, L., I. Kovrova, S. Lyon, M. Manacorda, and F.C. Rosati 2009. "Towards Consistency in Child Labour Measurement: Assessing the Comparability of Estimates Generated by Different Survey Instruments." Understanding Children's Work Project Draft Working Paper.

Hausman, J.A. 2001. "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the Left." *Journal of Economic Perspectives* 15 (4):57–67.

Hausman, J.A., J. Abrevaya, and F.M. Scott-Morton 1998. "Misclassification of the Dependent Variable in a Discrete Response Setting." *Journal of Econometrics* 87: 239–269.

Hill, D.H. 1987. "Response Errors in Labor Surveys: Comparisons of Self and Proxy Reports in the Survey of Income and Program Participation (SIPP)." In *Proceedings of the Bureau of Census, Third Annual Research Conference.*

Hussmanns, R., F. Mehran, and V. Verma 1990. *Surveys of Economically Active Population, Employment, Unemployment and Underemployment: An ILO Manual on Concepts and Methods.* ILO: Geneva.

Hyslop, D.R., and G.W. Imbens 2001. "Bias from Classical and Other Forms of Measurement Error." *Journal of Business & Economic Statistics* 19(4): 475–481.

ILO. 1982. Resolution Concerning Statistics of the Economically Active Population, Employment, Unemployment and Underemployment. Adopted by the Thirteenth International Conference of Labour Statisticians, Geneva: International Labour Organization.

Judge, G., and L. Schechter 2009. "Detecting Problems in Survey Data Using Benford's Law." *Journal of Human Resources* 44(1):1–24.

Kalton, G., and H. Schuman 1982. "The Effect of the Question on Survey Responses: A Review." *Journal of the Royal Statistical Society* 145(1):42–57.

Kasprzyk, D. 2005. "Measurement error in household surveys: sources and measurement." In *Household Sample Surveys in Developing and Transition Countries*, New York : United Nations, pp. 171–198.

Martin, E., and A.E. Polivka 1995. "Diagnostics for Redesigning Survey Questionnaires: Measuring Work in the Current Population Survey." *Public Opinion Quarterly* 59:547–567.

Martin, J., and B. Butcher 1982. "The Quality of Proxy Information: Some Results from a Large Scale Study." *The Statistician* 31:293–319.

Mata Greenwood, A. 2000. *Incorporating Gender Issues in Labour Statistics.* Geneva: International Labour Office, Bureau of Statistics.

Mathiowetz, N.A., and R.M. Groves 1985. "The Effects of Respondent Rules on Health Survey Reports." *American Journal of Public Health* 75(6):639–644.

Moore, J. 1988. "Self/Proxy Response Status and Survey Response Quality: A Review of the Literature." *Journal of Official Statistics* 4:155–172.

National Bureau of Statistics Tanzania. 2003. Integrated Labour Force Survey, 2000/01 – Analytical Report, see http://www.nbs.go.tz/labourforce/index.htm

———. 2002. Household Budget Report, 2000/01, Chapter 5. Dar es Salaam; http://www.tanzania.go.tz/hbs/Index_FinalReport_HBS.htm

Office for National Statistics (ONS). 2003. Labour Force Survey User Guide: Volume 1 Background and Methodology. United Kingdom.

Poterba, J.M., and L.H. Summers 1986. "Reporting Errors and Labor Market Dynamics." *Econometrica* 54(6):1319–1338.

——— 1995. "Unemployment Benefits and Labor Market Transitions: A Multinomial Logit Model with Errors in Classification." *Review of Economics and Statistics* 77(2): 207–216.

Sinclair, M.D., and J.L. Gastwirth 1998. "Estimates of the Errors in Classification in the Labour Force Survey and Their Effect on the Reported Unemployment Rate." *Survey methodology* 24(2): 157–169.

Singh, A.C., and J.N.K. Rao 1995. "On the Adjustment of Gross Flow Estimates for Classification Error with Application to Data from the Canadian Labour Force Survey." *Journal of the American Statistical Association* 90(430): 478–488.