WPS8147

Policy Research Working Paper 8147

Measuring Time Use in Development Settings

Greg Seymour
Hazel Malapit
Agnes Quisumbing

WORLD BANK GROUP
Africa Region
Office of the Chief Economist
&
Development Data Group
July 2017

Policy Research Working Paper 8147

Abstract

This paper discusses the challenges associated with collecting time-use data in developing countries. The paper suggests potential solutions, concentrating on the two most common time-use methods used in development settings: stylized questions and time diaries. The paper identifies a significant lack of rigorous empirical research comparing these methods in development settings, and begins to fill this gap by analyzing data from Women's Empowerment in Agriculture Index surveys in Bangladesh and Uganda. The surveys include stylized questions and time diary estimates for the same individual. The study finds limited evidence that stylized questions are more feasible (in terms of interview length) but also less accurate, compared with time diaries. These results are attributed to the relatively greater cognitive burden imposed on respondents by stylized questions. The paper discusses the importance of broadening the scope of time-use research to capture the quantity and quality of time, to achieve richer insights into gendered time-use patterns and trends. The paper suggests a path forward that combines mainstream time-use data collection methods with promising methodological innovations from other disciplines.

This paper is a product of the Gender Innovation Lab, Office of the Chief Economist, Africa Region and the the Living Standards Measurement Study, Development Data Group. It is part of a larger effort by the World Bank to provide open access to its research and make a contribution to development policy discussions around the world. Policy Research Working Papers are also posted on the Web at http://econ.worldbank.org. The authors may be contacted at G.Seymour@cgiar.org.

Measuring Time Use in Development Settings

Greg Seymour[1], Hazel Malapit[2], and Agnes Quisumbing[3]

Keywords: time use, recall error, measurement error
JEL codes: C8, J22, O12, Q12

[1] g.seymour@cgiar.org, International Food Policy Research Institute, CGIAR Research Program on Policies, Institutions, and Markets
[2] International Food Policy Research Institute, CGIAR Research Program on Agriculture for Nutrition and Health
[3] International Food Policy Research Institute, Poverty, Health, and Nutrition Division

1. INTRODUCTION

Time is increasingly recognized as a basic resource required for reducing poverty and improving well-being (Williams, Masuda, and Tallis 2015). A 2010 United Nations (UN) report finds that in all regions of the world: (1) women spend at least twice as much time as men on unpaid (household) work and (2) women's total work burden is higher than men's when both paid and unpaid work are taken into account (UNDESA 2010). Such patterns demonstrate the global persistence, particularly in less developed countries, of a traditional gender division of labor, in which women specialize in domestic responsibilities, such as the care of children and other family members (reproductive labor), and men specialize in activities that are more closely tied to income generation (productive labor). The unequal sharing of household work between the sexes imposes several costs on women's well-being and livelihoods. Notably, time spent on domestic work restricts women's access to paid employment and reduces their time available for education, leisure, self-care and social activities.

Designing effective policies or programs to address gender inequalities in time use is complex and requires considering the social context of men's and women's roles and responsibilities within the household among other factors. An important but often overlooked prerequisite to this process is the ability of researchers to accurately measure men's and women's time burdens and gender disparities in both productive and reproductive work. Various survey-based methods for measuring time use currently exist, such as stylized questions and time diaries, which have been applied in both developed and developing country settings. However, a lack of consistency across space and time in how these methods are used makes it difficult to compare results across countries, to capture changes in time use patterns over time, and to monitor progress in the unequal distribution of care work between men and women.

To advance our understanding of gendered time use patterns and trends in developing countries, this paper critically reviews methods used to measure time use. We begin by outlining common methods for measuring time use in developing countries. We follow this with a discussion of the challenges associated with measuring time use and a brief empirical analysis using data from Women's Empowerment in Agriculture Index (WEAI) surveys in Bangladesh and Uganda, which contain stylized

questions and time diary data for the same individual. We then discuss methods of data collection aimed at measuring quality of time (e.g., enjoyment, energy expenditure), and conclude with a broad discussion of the advantages and disadvantages of each of the methods considered in the paper and suggest ways in which elements of these methods might be combined to provide richer insights into gendered time use patterns and trends.

2. OVERVIEW OF TIME USE METHODS

Survey methods commonly used to collect time use data fall into two general categories: methods that focus only on specific activities (*stylized questions*) and methods that ask about all the activities undertaken within a specific period (*time diaries*).

Which method is best suited to collect time use data in developing countries? The answer largely depends on the research question of interest. In general, the types of research questions that can be answered by time use studies fall into two broad categories: questions that only require information on the absolute quantity of time spent on particular activities (unidimensional), and questions that require information not only on *quantity* but also on the *quality* of time spent on these activities (multidimensional). We consider two additional criteria when evaluating time use methods: to what degree does it provide estimates that match how people actually spent their time (accuracy) and how difficult is it to implement in a developing country setting, including cognitive concerns as well as time and financial costs (feasibility)?

2.1. STYLIZED QUESTIONS

Stylized questions typically ask respondents to estimate the actual or, in some cases, the "usual" or "typical" amount of time they devoted to a particular activity during a specific time interval (e.g., the previous day, week, month, or year).[4] They are commonly used in comprehensive household surveys

[4] The questions are "stylized" in the sense that they refer to a hypothetical construct (a "typical" day) rather than a specific time period (yesterday).

aimed at measuring time use over a specific set of activities—in which case a full accounting of the respondent's activity over the reference period would be needlessly cumbersome. A recent impact evaluation study in northwest Bangladesh provides an example (see Box 1).

Box 1. Stylized questions in an impact evaluation study of the CARE-Bangladesh Strengthening the Dairy Value Chain Project

Although the broad aim of the CARE-Bangladesh Strengthening the Dairy Value Chain Project was to improve the participation of smallholder farmers in the dairy value chain in northwest Bangladesh, several components of the project were specifically targeted to women, who are traditionally responsible for the care and feeding of livestock, even if these are regarded as mostly owned by men (Quisumbing et al. 2013). A key component of the project encouraged women to bring milk to collection points instead of selling them to itinerant buyers, which was expected to increase time that women spent on dairy-related activities. To assess whether the increased time involved in dairying had implications for domestic work, and whether increased time burdens were shared among household members, the authors asked respondents the following questions with respect 18 livestock-related activities and 5 household activities (see Figure 1):

- Who is primarily responsible for doing it? [MEMBER ID]
- Who does it in the absence of the responsible person? [MEMBER ID]
- In the past 30 days, who usually does it? [MEMBER ID]
- In an average week in the past 30 days, how many hours total has this person spent on the activity?

Importantly, the list of activities is not exhaustive. Instead, it focuses only on the specific areas of interest in which the project hypothesized that tradeoffs exist—livestock and household activities. Total time spent in these activities was aggregated for men, women, boys, and girls. The authors found that the increased allocation of adult women's time to dairy activities came at the expense of time in household activities, with young girls (but not boys) consequently increasing their time in domestic work.

Conventional wisdom suggests that stylized questions are the least costly of the methods considered in this paper, based on the argument that they collect much of the same data as more extensive methods (e.g., time diaries), but require far fewer questions and less effort (e.g., shorter interviews) to implement. This is clearly true in most developed country settings, where stylized questions are often used to collect a person's working hours in labor force surveys.[5] It is much less obvious whether stylized

[5] See, for example, the labor force component of the United States Current Population Survey: http://www2.census.gov/programs-surveys/cps/techdocs/questionnaires/Labor%20Force.pdf.

questions are the most feasible option in development settings, where the cognitive requirements involved may be difficult to cope with for less-educated enumerators and respondents (see discussion below).

2.2. TIME DIARIES

Rather than focusing on a single activity as is conventionally followed with stylized questions, *time diaries* collect information about all of a respondent's activities during a specific reference period, typically the previous 24 hours. They can be self-reported or completed with the assistance of an enumerator.

With self-reported time diaries, respondents record all of their activities during the reference period, typically in 15 or 30 minute increments. Self-reporting can be done contemporaneously using a paper diary, or more recently an electronic device (e.g., tablet, smartphone), supplied to them by the research team. It can also occur retroactively via phone- or internet-based surveys, depending on the availability of technology. Self-reported time diaries require a high level of literacy, numeracy, and (in some cases) technological savvy, and as a result, are relatively untested in developing countries, though a few exceptions exist (e.g., Floro and Pichetpongsa 2010; Masuda et al. 2014).

Survey-based time diaries—by far the most common type used in developing countries—are completed retroactively with the help of an enumerator during a standard household survey interview based on respondents' recollection of their activities during the reference period (again, usually the previous 24 hours). The actual structure of the interview can vary depending on the expertise of the survey team. Most commonly, respondents are asked to describe the sequence and duration of their activities over the reference period in chronological order, e.g., beginning with the time they woke up and ending when they went to sleep. See Box 2 for an example of a survey-based time diary from the WEAI (Alkire et al. 2013).

Box 2. Time diary in the Women's Empowerment in Agriculture Index survey

The WEAI survey collects time use data for both primary and secondary (simultaneous) activities across the following 18 types of activities: [6]

- Personal activities: sleeping and resting, eating and drinking, personal care, school (including homework)
- Market work: work as employed, own business work, farming/livestock/fishing
- Unpaid (household) work: shopping/getting service, weaving/sewing/textile care, cooking, domestic work (including fetching wood and water), care for children/adults/elderly, traveling/commuting
- Leisure: watching TV/listening to radio/reading, exercising, social activities and hobbies, religious activities

Originally designed to be used as a paper survey, the WEAI time diary follows a format adapted from the Lesotho Household Budget Survey. In this format, rows representing each type of activity and columns representing the previous 24 hours divided into 15-minute intervals combine to form an 18 × 96 grid (see Figure 2).[7] Activities are recorded by drawing horizontal lines across the grid. In this way, the format itself acts to promote the accurate recording of data. By simply checking that a continuous (though, of course, staggered) line extends across the grid, enumerators are able to verify that all of the previous 24 hours have been counted and that no additional time has been recorded in the module. Indeed, errors of this sort are rare in data collected using the WEAI time diary; though, of course, without a proper comparison group we cannot say for certain whether this is truly the case. Moreover, the possibility exists that the grid format may be too complicated for some enumerators to handle, as reported by some implementers of the WEAI. As such, extra training may be required to prepare enumerators for the WEAI time diary.

More recently, the WEAI time diary has been adapted for use with computer-assisted personal interviewing (CAPI) software on a tablet or smartphone, and the grid format discarded in favor of a set of nested, "drop-down" menus that allow for greater specificity of listed activities, while also providing built-in checks for data consistency. The module is currently being piloted in several developing countries.

Survey-based time diaries are, at present, the most common method for collecting time use data in developing countries, having been implemented in at least 57 such countries around the world (Fisher 2015). The method is widely perceived to be both cost-effective to implement (albeit more costly than stylized questions) and to provide accurate estimates of how people spend their time. Hence, many researchers consider the survey-based time diary to be the "gold standard" for collecting time use data in developing countries.

[6] An "Other (specify)" option is also included in the module to capture any activity that does not fall into one of the listed categories.
[7] See Lawson (2007) for details on the Lesotho Household Budget Survey time use module.

A close reading of the time use literature, however, fails to uncover sufficient justification for this claim. In terms of feasibility, the type of data needed to evaluate the cost-effectiveness of survey-based time diaries versus other methods (e.g., survey costs, interview length) are rarely shared by organizations. Equally rare are follow-up surveys (e.g., qualitative interviews, cognitive testing) aimed at evaluating respondents' experience with different methods—for instance, how easy was it for the respondent to recall a certain activity? Where comparative data on the feasibility of time use methods does not seem to exist at all, data abound on the comparative accuracy of time use methods, particularly the question of stylized questions versus time diaries (Juster, Ono, and Stafford 2003; Kan and Pudney 2008; Frazis and Stewart 2012). This research generally tends to favor time diaries, noting the presence of greater systematic error in estimates based on stylized questions. However, these studies almost entirely rely on data from developed countries, mostly from the United States or United Kingdom. Virtually no rigorous empirical research exists comparing these methods in development settings.

3. COGNITIVE ASPECTS OF TIME USE METHODS

To understand the specific challenges involved in collecting time use data in development settings we draw on insights into the cognitive aspects of survey methods from the psychology literature.[8] Consider, for example, the following question from the WEAI 2.0 pilot survey: "In the last seven days, how much time did you spend farming?" To accurately answer this question respondents must: (1) properly identify the relevant reference period ("the last seven days") and the specific activities in question ("farming"); (2) search their memory of the past seven days to retrieve all instances of farming activities, and (3) correctly aggregate all of these into a frequency estimate. Each one of these steps presents unique challenges to respondents.

Perhaps the most common of these challenges relates to the ability to correctly identify the activity in question. This problem often has to do with how questions are phrased. Broad terms, such as

[8] Although this evidence comes almost entirely from subjects in developed countries, we believe many of the insights are likely to be relevant in developing countries.

"farming" in the example above, can be interpreting in many different ways. For example, we included a category for "farming" in the original time diary used in the WEAI survey, with the intention that it be used to capture any crop- or livestock-related activity. However, this guidance was not effectively communicated to the survey team in Bangladesh, who instead tended to interpret "farming" as only referring to agricultural activities performed away from the homestead, such as cultivating field crops. As a result, all activities performed by women on the homestead, regardless of whether they were related to agriculture, were classified as "domestic work." This resulted in the misclassification of time spent tending livestock on the homestead, as well as home gardens. As a result, estimates of "farming" based on this data severely underestimated the true extent of women's agricultural work, and because other non-farming tasks were also classified as "domestic work," severely limited the type of research questions the data could reliably answer. Guidance documents were quickly revised with instructions designed to prevent this mistake from being repeated in other WEAI surveys—enumerators are now instructed to provide respondents lists of specific activities typically associated with farming in the study setting, regardless of where the activity took place. Nevertheless, the lesson from our experience with the WEAI is relevant to researchers interested in measuring time use. Time diaries should be designed to ensure a consistent understanding across respondents (and enumerators) about which specific activities fall within each category. Otherwise, the time use estimates produced from the data may be meaningless for interpersonal comparisons. While a certain degree of ambiguity may, in rare cases, be desirable—if, for instance, a person's perception of what constitutes an activity is important—in most cases, it is best avoided in favor of precise language or through the inclusion of locally-relevant examples.

Confusion can also arise from problems having to do with the ability to correctly identify the reference period. Similar to the previous example, such risks can be partially mitigated by avoiding ambiguity in how questions are phrased. For example, asking about a fixed reference period, such as "the last seven days," is much more likely to yield consistent interpretation, than asking about vaguely defined or hypothetical reference periods (e.g., last several weeks, usual day).

However, even if respondents are able to identify the intended reference period, they may still encounter problems when it comes to accurately assigning memories to the reference period. This is because autobiographical memory tends to be organized in the context of ongoing life experiences, rather than in relation to specific dates (Belli 1998). Thus, temporal-based attempts at anchoring—common phrases used such as "in the last month" or the "during the last 12 months"—may be relatively ineffective. Instead, prompting respondents with locally-relevant recall cues (e.g., public events, religious services) or graphical aids (e.g., personal event timelines) to help anchor the reference period to respondents' personal experiences (Loftus and Marburger 1983; van der Vaart and Glasner 2007). The use of shorter recall periods (last week vs. during the last 12 months) also increases the chances that respondents will be able to recall specific activities that occurred during the reference period, whereas longer recall periods tend to encourage guessing and estimation based on established patterns of activity (Brown 2002). Even then, respondents may exhibit "telescoping" behavior, particularly at longer recall periods, in which distant events may be reported as having happened more recently than they did ("forward telescoping") and recent events as occurring much longer ago ("backward telescoping") (Sudman and Bradburn 1973).

It is also important to stage the time use interview in a way that complements, rather than competes against, the chronological ordering of autobiographical memory. For example, asking respondents to recall when and for how long they engaged in each activity listed in the time diary may be particularly challenging for respondents, whereas encouraging the respondent to recount the previous day as narrative account may, in fact, enhance recall accuracy.

Guidance for the WEAI time diary provides an example of one way to structure time diary interviews to promote better recall (see Box 3 for another example from the United States).[9] The WEAI time diary interview begins with the enumerator asking the respondent to recall what time he or she woke

[9] These guidelines were developed based on the protocol developed for the WEAI by Data Analysis and Technical Assistance Limited (DATA) in Bangladesh. A video tutorial for the method is available online: https://www.youtube.com/watch?v=jr8ebiKUkbQ.

up and went to sleep the previous day; this establishes the boundaries for the period of time that must be "filled in" during the interview. Next, the enumerator asks the respondent what they did immediately after waking up and for how long. The interview proceeds in this manner, from one event to the next, until a full account of the respondent's day is captured. The enumerator's primary task throughout the interview is to "translate" the respondent's narrative of the previous day (what he or she did and for how long) into the coding scheme of the time diary, only interrupting for clarification when absolutely necessary. For instance, if the respondent reports spending time tending crops in the field, the enumerator seamlessly codes this as "farming" without interrupting the narrative. The goal of the interview is to reconstruct the previous day organically, proceeding from one event to the next according to the respondent's internal chronology of the previous day with minimal prompting from the enumerator.

Box 3. The Day Reconstruction Method

Kahneman et al.'s (2004) Day Reconstruction Method (DRM) asks respondents to recall their activities from the previous day using a structured self-administered questionnaire. In the DRM, respondents are asked to write a narrative description of the previous day, in which they describe each episode of activity. Following this, they are asked a series of follow-up questions about each episode (e.g., what they were doing, how long they were doing it for, where they were, who they interacted with, and how they felt). By prompting respondents to spend a few minutes reflecting on the previous day, the DRM attempts to stimulate respondents' memories of the previous day prior to actual data collection.

As a self-administered questionnaire, the DRM is likely ill-suited for use in developing countries due to its steep cognitive requirements. However, an abbreviated version of the DRM, designed by researchers at the World Health Organization (WHO), has been implemented in several developing countries (Miret et al. 2012; Ayuso-Mateos et al. 2013). The abbreviated DRM, however, lacks several of the innovations that made the DRM such an innovative tool for measuring time use and for reducing the risk of recall bias. Namely, it does not call for a pre-interview reflection period by respondents—potentially a key component for encouraging accurate recall—and only captures a portion of the previous day.

Another cognitive aspect of time use methods (closely related, but separate from the ability to accurately recall events during the reference period) that may complicate measurement concerns the ability to accurately aggregate specific instances of activities over the reference period. Typically, psychologists believe that a person's ability to accurately formulate frequency estimates is directly related to how often an activity occurs (regularity) and how distinguishable it is from other activities (saliency)

(Menon 1993; Brown et al. 2007). Frequency estimates for activities that occur regularly but are not exceptionally salient (e.g., eating breakfast, brushing your teeth) are likely to be extrapolated based on the respondents' assessment of the rate of occurrence. More salient activities, on the other hand, are likely to evoke a "recall-and-count" strategy, in which respondents recall each individual occurrence of the activity and aggregate them across the reference period. Neither estimation strategy is ideal, however, when activities do not follow a regular schedule and are not especially salient, such as may be the case with agricultural labor. As a result, stylized questions about the amount of time a person spends on these types of activities are likely to yield particularly erroneous estimates. This is consistent with the findings from empirical research investigating recall bias in agricultural labor statistics (Arthi et al. 2016) and time use evidence from developed countries (Juster, Ono, and Stafford 2003): stylized questions tend to yield less accurate estimates than time diaries for irregularly occurring activities, and conversely, more accurate estimates for regularly occurring activities.[10] As a result, stylized questions may be better suited for answering research questions about activities that tend to follow a set schedule (e.g., salaried employment), where the risk of recall bias may be considerably less.

3.1. SEASONALITY

Farming is an inherently unpredictable occupation. The amount of time a person spends farming is likely to exhibit wide swings not only from one season to another but also within a season, depending on myriad factors (e.g., weather, crop choice, technology used).[11] Even beyond the cognitive factors discussed above, what makes measuring time use in developing countries so challenging is that so many people rely on agriculture as their primary livelihood strategy.

[10] Similarly, Kan and Pudney (2008) compare stylized question and time diary estimates for the same individual in a British survey and find evidence of greater measurement error in stylized questions versus time diaries, though they attribute this mostly to randomness rather than systematic bias.

[11] Seasonality bias is not limited to farming. Time spent on other activities that also follow a seasonal schedule (e.g., small businesses, construction work, migrant labor) may also exhibit seasonality bias.

In contrast to the cognitive issues discussed above, which tend to originate from *how* time use questions are asked, seasonality bias, as it is commonly referred to, stems from *when* time use questions are asked. Depending on the length of the reference period, time use surveys typically only capture a small cross-section of how a person spends their time over the course of a year. The shorter the reference period, the more likely it is to miss seasonal variation in time use patterns. Thus, seasonality bias is a much greater concern for time diaries, which tend to focus on a single 24-hour period of time, than for stylized questions, which may cover multiple days—though, of course, as noted above, there are other risks associated with using longer reference periods. While there are certain precautions that can be taken, such as follow-up questions designed to distinguish between patterns of time allocation that are out of the ordinary (e.g., holidays, festivals, etc.) or posing questions about "usual" or "typical" days, none of these are perfect solutions. Ideally, time use data should be collected at multiple times for each respondent to ensure coverage across seasons (Frazis and Stewart 2012). However, given concerns over rising costs, such a sampling strategy is unlikely to be tenable for most researchers. In this case, the best advice may be to become informed about the seasonal labor patterns associated with the study setting, perhaps by conducting qualitative interviews or focus group discussions prior to the survey, and to stage time use data collection at the most appropriate time for your particular research question.

4. FEASIBILITY AND ACCURACY OF STYLIZED QUESTIONS AND TIME DIARIES

4.1. FEASIBILITY

Of the two most common time use methods used in developing countries, stylized questions and time diaries, which is the most feasible for use in a developing country setting? It is possible that despite the outward appearance of fewer costs (e.g., fewer questions added to the survey), stylized questions may, in fact, be costlier to implement than time diaries, when one takes into consideration potential increases in the cognitive burden imposed on respondents, which may produce longer interview times. To answer this question, we look to data from the pilot testing of the WEAI 2.0 survey in Bangladesh and Uganda in 2014.

As noted above, the WEAI time diary collects information about the time men and women spent on a wide range of work and non-work activities during the previous 24 hours. Strictly speaking, however, the WEAI only requires information on the time men and women spend on market and non-market work activities, which we use to assess whether a person has an excessive workload (see Alkire et al. 2013 for details). In response to concerns raised by implementers about the time required to conduct the WEAI time diary interview, we designed a stylized question version of the WEAI time diary, aimed at "streamlining" time use data collection in the WEAI. The questionnaire asks respondents the following questions with respect to 12 types of market and non-market work (see Figure 3):

1. In the last seven days, how much time in hours did you spend on [ACTIVITY]?
2. Did you spend a usual amount of time on [ACTIVITY] in the last seven days?
3. Since the last week was not usual, within the last six months how much time did you usually spend on [ACTIVITY] per week?[12]

We expected that the stylized questions would be easier for respondents and enumerators to understand and require less time to implement in the field compared to the time diary, which collects information on a wider range of activities. A survey experiment designed to test this hypothesis was included as part of the pilot testing of the WEAI 2.0 survey in Bangladesh and Uganda in 2014.

Sample villages were randomly assigned to either the original WEAI time diary with primary and secondary activities, or an experimental time use module, including stylized questions plus a time diary collecting only primary activities. For respondents assigned to the latter group, the stylized questions were asked directly prior to the time diary during the same visit. Enumerators were instructed to record the time at the beginning and end of each survey module. Basic descriptive statistics for the data are given in **Table *1.* Descriptive statistics of WEAI 2.0 pilot survey by country**Table 1.

[12] Question 3 was only asked if the time spent on [ACTIVITY] in the last seven days was unusual according to the respondent.

Table 2, we compare the length of time required to complete the survey module for each method. Although in Bangladesh the stylized questions did, in fact, take less time to complete than the time diary (by roughly two minutes), in Uganda we observed no significant difference in interview length between the two methods. Thus, contrary to our expectation, we found that stylized questions did not always produce shorter interviews than time diaries. We should note, however, that the sequencing of the time diary immediately following the stylized questions may be problematic for comparing interview lengths. On one hand, the stylized questions and time diary purposely utilized different recall periods (7-day vs. 24-hours, respectively), which should minimize the risk of contamination. On the other hand, it is possible that respondents' ability to recall their activities during the past 24 hours was, in some way, impacted by first asking the stylized questions. Even then, it is unclear in which direction the time diary interview lengths would be biased: downward (if the stylized questions positively stimulated respondents' recall of the past 24-hours) or upward (if after answering the stylized questions respondents had difficulty restricting their focus to the past 24-hours).

Qualitative evidence from the field, however, provides a convincing explanation for why the stylized questions did not always produce shorter interviews. According to the survey teams in both countries, respondents often found the stylized questions to be more challenging than the time diaries—perhaps owing to the cognitive challenges discussed above. Indeed, many enumerators reported that they had to help respondents to extrapolate a weekly total based on whatever the respondent could remember—even if this was incomplete—which suggests that the stylized questions estimates may contain a high degree of measurement error. Field reports from both countries also revealed that the stylized questions required more training time than the time diary. However, given that many enumerators had participated in the first round of WEAI piloting in 2011—and hence, were familiar with the format—but lacked experience with the stylized questionnaire, it is difficult to say for certain whether stylized questions would require more training time than the time diaries in other situations. Indeed, we suspect that stylized questions would, in fact, require less training time than the time diaries in most cases. Nevertheless, the results of our experiment reveal little reason to believe that stylized questions can

reliably produce shorter interview times than time diaries. As such, current iterations of the WEAI continue to utilize the time diary format.

The ultimate decision of which method to use, however, requires consideration of factors beyond simply which method is most feasible. Researchers much consider which method makes the most sense for their research question. One component of this decision is the level of accuracy required to answer the research question. For instance, understanding how individual time allocation patterns evolve over time would require greater accuracy than comparing average time allocation patterns across gender at a single point in time. Hence, in the next section, we consider the relative accuracy of stylized questions and time diaries.

4.2. ACCURACY

One of the fundamental goals of time use research is to establish patterns of individual time allocation. Which activities take up the most time in a person's day? How much time does a person save by using a new water collection system? Does participation in a particular agricultural activity reduce women's time available for the care and feeding of young children? Investigating questions like these requires accurate information about the amount of time people spend on different activities. As the previous discussion has noted, however, the complexities involved in answering time use questions mean that respondents' own estimates of how they spend their time may not be accurate and, moreover, that the extent of these inaccuracies may vary depending on the activity in question and survey method.

In this section, we compare time use estimates obtained from the same individuals using two different methods—time diaries and stylized questions and attempt to explain these differences in terms of several potential factors. As before, the data come from the WEAI 2.0 pilot survey in Bangladesh and Uganda.

Admittedly, our method is imperfect; the WEAI pilot survey was not specifically designed with this type of analysis in mind. The primary problem owes to differences in recall periods. The time diary asks about the previous 24 hours; the stylized questions ask about the past seven days. While the average

time spent on an activity during the past seven days may, in fact, correspond to the time spent on that activity during the past 24 hours, there is no reason to believe that this should always be the case. In fact, there are several reasons why we might expect this not to be true (e.g., sickness, weather shocks). Unfortunately, given the design of the survey, there is no way to determine how much of the observed difference between methods can be attributed to the conflicting recall periods and how much stems from recall error or other reporting biases. Nevertheless, we believe that there are valuable insights to be gained from undertaking this analysis, despite the flaws in the underlying data. Indeed, given the dearth of rigorous empirical research comparing time use methods in development settings, even analysis based on imperfect data is a step forward.

The first question we seek to answer concerns whether differences in time use estimates vary systematically according to the type of activity. This could be the case, if, for instance, certain activities are more difficult to recall than others. Based on the discussion in the previous section, we expect the differences between the two methods to be greatest for irregular, less salient activities, such as agricultural work.

We begin our analysis by estimating country-specific OLS regressions that examine whether differences between estimates based on time diaries and stylized questions vary systematically according to the activity. The dependent variable is the difference in the respondents' stylized and time diary estimates at the activity-level, and the explanatory variables comprise a set of dummy variables corresponding to each type of activity. If a certain activity is more prone than others to recall error or other biases, then the dummy variable should be statistically significant. Since the unit of analysis in this regression is at the activity-level, each respondent may be represented in the sample up to nine times (the number of activities asked about). To account for this fact and for the interconnected nature of time allocation patterns for individuals within the same household, we cluster standard errors at the household-level.[13]

[13] An alternative approach would be to cluster standard errors at the individual-level. Our results are robust to either specification.

Comparison of the average time spent in each activity and the statistical significance of the difference (based on paired *t*-tests) for respondents in both countries (Table 3) and regression results (Table 4) reveal both activity- and country-specific patterns. Respondents report spending roughly the same amount of time on employed labor regardless of the method, but reported time spent on shopping and receiving service, care work, and traveling/commuting differs significantly depending on the method. These patterns persist in both countries, and are broadly consistent with the notion that a respondent's ability to accurately recall an activity is affected by the regularity and saliency of the activity. Time spent on employed labor is likely to follow a very regular schedule; hence, respondents' ability to accurately report their time tends not to be greatly affected by the longer recall period associated with stylized questions.

The difference in reported time use in other activities, however, is not consistent across countries, nor does it fully conform to the predictions of the cognitive psychology literature. This is particularly true for time spent on farming, livestock, and fishing activities. Given that agricultural tasks tend to follow an irregular schedule and can often be monotonous (low saliency), we expected to find significant divergence between stylized estimates and time diary estimates—based on the premise that respondents would struggle to accurately recall time spent on agricultural work over a 7-day recall period. While we do find evidence of such a trend in Uganda—time spent on agricultural activities is 83 minutes higher in stylized estimates relative to time diary estimates—respondents in Bangladesh report spending roughly the same amount of time on these activities regardless of question type. Similar trends can be observed for own business work, weaving/sewing/textile care, cooking, and domestic work.

Overall, the results suggest that stylized questions and time diaries can, in certain settings, provide equally accurate estimates, even when it comes to highly irregular, agricultural activities. Yet, we would advise not reading too much into the country-specific differences we observed, given that they could be driven by any number of unobserved factors. For instance, it may be that more experienced enumerators are able to extract more accurate estimates from respondents regardless of the type question used. In fact, in Bangladesh, where enumerators tended to be highly qualified, often holding an advanced,

post-secondary degree, differences between the two methods tend to be quite low regardless of activity type: the largest average discrepancy observed in the Bangladeshi data is 13 minutes (traveling or commuting), compared to 83 minutes in the Ugandan sample (farming, livestock, or fishing activities).

Next, we investigate potential sources of reporting bias that could impact differences in the time use estimates provided by the two methods. Reporting bias may arise for several reasons (see the discussion in the previous section). One of the most common forms is social desirability, which occurs when respondents attempt to (internally) edit their responses to the time use questions before they communicate them to enumerators due to the influence of factors, having to do with how respondents want to be perceived by enumerators and/or others in their community. We test for the presence of reporting bias a similar empirical framework and the same data as before. Our dependent variable is, again, the difference between respondents' time diary and stylized question estimates. Models 2, 3, and 4, however, introduce additional sets of explanatory variables capturing three difference potential sources of reporting bias: respondent characteristics (sex, age, education, literacy); enumerator fixed-effects (i.e., dummy variables associated with each enumerator); and whether the respondent was interviewed alone or with other adults present.

Table 5 shows our results. Though not always consistent across countries, we find evidence of effects from respondent characteristics and enumerator fixed-effects, but not for whether the respondent was interviewed alone. Sex and education are statistically significantly related to differences between time diary and stylized question estimates for respondents in Uganda, though not in Bangladesh. Possible explanations for this pattern may stem from differences in the types of activities traditionally undertaken by women in the two countries, with women being more involved in a wider range of productive agricultural activities in Uganda and being less subject to barriers to mobility compared to Bangladesh, as well as the possibility that activities in monoculture rice cultivation (as in Bangladesh) may not vary much within the short span of time referred to in the interview. Enumerator fixed-effects, on the other hand, are jointly significant in both countries. This suggests that there may, indeed, be some correlation

between respondents' ability to accurately recall their time use and enumerators' abilities, and speaks to the importance of proper enumerator training.

5. QUALITY OF TIME

Traditionally, time use research has focused on the *quantity* of time. Research on *quality* of time research moves beyond this, and attempts to understand not only how much time people devote to different activities but also how they experience time. For instance, how do people's emotional states change as they move from one task to another over the course of a day? Some tasks may bring pleasure or happiness, whereas other tasks may be more unpleasant or physically taxing, due to higher energy expenditure or some element of danger or risk. Information about quality of time might also shed light on how labor burdens are shared within households, by making comparisons between individuals within the same household more meaningful. Indeed, women often perform multiple activities at once (e.g., livelihood activities and childcare), which can be more stressful or demanding than engaging in a single activity.

Accurately measuring the quality of a person's time is, however, a difficult proposition. One approach is to draw inferences about a person's physical or emotional state based on the activities he or she engages in, though this requires strong assumptions on the part of researchers. For example, Floro (1995) and Floro and Pichetpongsa (2010) draw conclusions about the deleterious effects of overlapping work activities on women's well-being, based on assumptions about the physical and emotional demands of combining work activities. Similar assumptions are implicit in the workload indicator of the WEAI, which is based on the total amount of time a person spends working: time spent on secondary work activities is given half the weight of primary work activities, regardless of the type or sequence of activities being combined.

5.1. EXPERIENCE SAMPLING

Assumptions such as these, however, can be problematic, given the idiosyncrasies involved in how people experience time (Krueger et al. 2009). For example, not all leisure activities are equally enjoyable to everyone, and not all non-leisure activities are equally arduous. Indeed, the most common approach to measuring psychological aspects of quality of time—broadly defined as *experience sampling*—avoids such assumptions and relies instead on respondents' own evaluations of their physical or emotional state.[14] It includes methods like the Experience Sampling Method (ESM) (Hektner et al. 2007), which asks respondents to record specific details about the activities they are currently engaged in at random moments throughout their day, usually prompted by a preprogrammed device (e.g., stopwatch, pager, timer, or smartphone), as well as several other adaptations of the ESM.

Although interest in experience sampling has mostly (and, perhaps, understandably) been limited to psychologists, we believe there are several promising avenues for methodological innovation based on experience sampling that may be worth considering for social scientists interested in time use research (see Box 4 for an example). Much of this has to do with the potential to eliminate the risk of recall bias. Assuming full compliance on the part of respondents and the absence of incentives to misrepresent what one is actually doing (e.g., if the respondent is engaged in illicit or illegal activities), experience sampling—recorded contemporaneously rather than retroactively—should yield exact data on respondents' activities.

Box 4. Insight into the intrahousehold sharing of labor in Bangladesh

Seymour and Floro (2016) successfully incorporated aspects of the DRM with a traditional, 24-recall time diaries to yield insight into the intrahousehold sharing of labor in Bangladesh. In their survey, respondents were asked to report how often they experienced five emotions (happiness, sadness, tiredness, pain, and stress) during each episode of activity indicated in the time diary on a 10-point Likert scale, ranging from 1 ("did not experience the feeling at all") to 10 ("experienced the feeling all

[14] We call this method "experience sampling," although it belongs to a family of methods that is much broader. Also called "event sampling methodology," this method was initially developed by Larson and Csikszentmihalyi (1983). Antecedents of this method include random spot observation. The crucial difference is that in random spot observation, external observers collect the data, whereas in experience sampling, the respondent responds about his or her conditions at a particular point in time.

the time"). Episodes were classified as unpleasant or pleasant depending on the most intense emotion experienced during the episode.[15]

Examining the proportion of time men and women in the sample experienced as pleasant for several different categories of activities (see Figure 4) provides insight into the intrahousehold sharing of labor among couples in Bangladesh. Although men and women reported roughly the same amount of overall working time (considering both labor market and household activities), men allocated more of their time to labor market work (75% of total work), whereas women allocated more time to household work (86% of total work). This gender gap is made even more meaningful when we consider how men and women experienced time spent in these activities. Women in the sample experienced 65% of time spent on household work as pleasant, compared to 84% for men. Women's negative experience is largely attributable to a few specific domestic activities (cooking, cleaning the home, and collecting water and firewood) rather than with care work, which women in the sample tend to think of positively (92% of time spent on care work is experienced as pleasant compared to 64% for non-care household work). Labor market work, on other hand, was experienced as pleasant at roughly the same rate by men and women in the sample.

Reticence among time use researchers to use experience sampling in developing countries is likely due to the impression that experience sampling methods are prohibitively costly to implement, given that respondents must be equipped with appropriate data collection devices. In the current global environment of decreasing technology costs and high levels of mobile phone usage, however, these constraints are quickly becoming less binding. We believe that one of the keys to improving current time use methods lies in technological innovation based on experience sampling. For instance, wearable activity-tracking devices used to track the movement of individual respondents and yield richer datasets, capable of answering a wider set of research questions.

5.2. ENERGY EXPENDITURE

Aside from the psychological aspects of time use, there is growing interest in monitoring another aspect of quality of time as well: energy expenditure or work effort associated with various activities. Drawing from efforts to measure physical activity, the key concepts revolve around frequency, duration, and intensity (FAO 2004). Development practitioners could be concerned, for example, with whether an intervention "did no harm" in terms of increasing women's energy expenditure. This question can be

[15] For example, if a respondent reported experiencing happiness (10) more often than sadness (2), tiredness (4), pain (1), and stress (3) during a particular episode of activity, then the episode was classified as pleasant.

broken down into three components: (1) whether the intervention increased the duration, intensity, and frequency of engagement in intervention activities; (2) if yes, whether it led to any adverse changes in BMI (e.g., decreasing weights may indicate that energy intakes do not compensate for greater activity levels); and (3) whether the intervention caused any other undesirable effects on time, such as reduction in time for child care, relaxation, or leisure.[16] These questions are especially relevant in developing country settings where lack of publicly provided infrastructure, limited access to mechanical or animal power, or access to labor-saving devices is associated with higher work burdens and increased drudgery. This approach is quite different from the developed country literature, where the emphasis is on increasing activity levels to prevent overweight and obesity, rather than to save labor and reduce drudgery.

Although the technology exists to measure energy expenditure and physical activity, it is only recently that advances have been made that reduce costs of monitoring energy expenditure in a field setting. The "gold standard" for measuring energy expenditure, in which the individual drinks doubly labeled water with heavy isotopes, and decreases in their concentration are measured in the person's urine, is costly and therefore infeasible to collect in large enough samples in a field setting, although studies have been conducted among smaller samples in developing countries.[17] Moreover, to get a picture of energy expenditure over an extended period of time—a period that could potentially affect weight or BMI—one should be able to measure the intensity (whether moderate or vigorous), duration (minutes or hours), and frequency (number of days per week) of activity. Several methods exist for obtaining this information (e.g., heart rate monitoring, pedometers or activity-tracking devices), but the most common is activity (or exercise) diaries. Activity diaries face similar issues as time diaries (for recording the quantity

[16] This draws from notes prepared by Gina Kennedy on measuring energy expenditure and physical activity, see http://www.a4nh.cgiar.org/files/2015/01/Energy-Expenditure-and-Physical-Activity-Reference-Notes1.pdf.

[17] A recent example of this approach in a developing country context is the study by Pontzer et al. (2012) in Northern Tanzania where they measured the total daily energy expenditure of 30 Hadza adults over an 11-day period.

of time): once the amount of time spent on a specific activity is determined (which is, of course, subject to recall bias), the activity has to be matched to a list of activities in various categories (e.g., occupational, home, leisure/recreation, travel or commuting), each with its own energy expenditure conversion factors.[18] Because respondents are unlikely to recall whether such activities were consistently performed at peak intensity or involved rest periods, estimates of energy expenditures using these methods are likely to be fraught with error.

6. CONCLUSION

In this paper, we discuss the challenges associated with collecting time use data in developing countries and suggest potential solutions, concentrating for the most part on the two most common time use methods used in development settings: stylized questions and time diaries. We identify a significant lack of rigorous empirical research on the accuracy and feasibility of these methods, and begin to fill this gap using data from WEAI surveys in Bangladesh and Uganda, which collected stylized questions and time diary data for the same individual. We find that stylized questions are sometimes less efficient (in terms of interview length) than time diaries—a result which we attribute, based on qualitative evidence, to the greater cognitive difficulties faced by respondents when answering stylized questions compared to time diaries. We also find evidence of significant differences in the time use estimates provided by stylized questions and time diaries, which can be linked to enumerator fixed-effects (in both countries) and respondent characteristics (in Uganda alone). The evidence of significant enumerator fixed-effects, in particular, should be carefully noted: if enumerators' abilities help to determine respondents' ability to accurately recall their time use, then proper enumerator training is paramount. Further research, at a greater scale than offered here, is needed to determine the broader validity of our results.

Considered together, no one method stands out as universally superior to the others. Each has its strengths and weaknesses. Traditional methods that account solely for the quantity of time still have to

[18] There are several questionnaires, mainly from the US and Europe that include links to physical activity questionnaires, for example: http://appliedresearch.cancer.gov/resource/collection.html.

address problems of recall bias, particularly in low-literacy populations, as well as the possibility of obtaining an inaccurate picture of time use patterns owing to seasonality. Methods that recognize multiple aspects of time use have greater potential for measuring well-being. In developing country settings, in particular, the energy intensity of activities, the increased work intensity owing to multi-tasking of productive and domestic activities, and the status and vulnerability associated with different types of work all combine to make the measurement of time a more nuanced effort than accounting for quantity alone. Addressing these issues in a cost-effective manner is essential if time use is to be captured in large-scale surveys that go beyond laboratory or small experimental studies.

We believe that broadening the scope of time use research to capture both quantity and quality of time use will provide richer insights into gendered time use patterns and trends, and that the best path to this goal will be through combining mainstream time use data collection methods with promising methodological innovations from other disciplines. Lessons from psychological approaches, in particular, are useful in discerning whether activities contribute to a person's greater sense of well-being. Approaches from the biomedical sciences can arrive at reasonable estimates of energy expenditure, which is important in determining the nutrition and health status of poor populations. As biomedical sensors become cheaper to use in field settings, such efforts may be useful in informing the development of technologies that reduce drudgery, particularly for women in their domestic roles. Careful documentation and comparison of these approaches will help to advance the multidimensional measurement of time, in order to identify what time constraints—whether actual or multidimensional—contribute most to gender inequality, and what policies and interventions can be implemented to relieve those constraints.

Table 1. Descriptive statistics of WEAI 2.0 pilot survey by country

	Bangladesh		Uganda	
	Mean	S.D.	Mean	S.D.
Female	0.551	0.498	0.558	0.497
Age	43.3	14.0	45.8	17.7
Highest level of education completed				
Less than primary	0.544	0.499	0.625	0.485
Primary	0.093	0.290	0.168	0.374
Secondary or above	0.361	0.481	0.207	0.406
Literate	0.551	0.498	0.582	0.494
Interviewed alone	0.744	0.437	0.970	0.172
Observations	399		328	

Source: Authors' calculations based on WEAI 2.0 pilot data.

Table 2. Average interview length by method and country (in minutes)

Country	24-hour (TD)	7-day (SQ)	Difference (TD - SQ)
Bangladesh	12.88	10.63	2.25***
Uganda	7.86	7.97	-0.11

Source: Authors' calculations based on WEAI 2.0 pilot data.
Notes: N=356 for Bangladesh; N=315 for Uganda. Paired *t*-tests performed comparing the mean survey times between methods. TD = time diary. SQ = stylized question. ***, **, and * denote statistical significance levels at 1%, 5%, and 10%, respectively.

Table 3. Average time spent on different activities by method and country (in minutes)

	Bangladesh			Uganda		
Activity	24-hour (TD)	7-day (SQ)	Difference (TD - SQ)	24-hour (TD)	7-day (SQ)	Difference (TD - SQ)
Work as employed	35.49	44.25	-8.76	12.53	14.09	-1.55
Own business work	39.96	30.96	9.00*	33.57	48.03	-14.46**
Farming/livestock/fishing	152.86	143.26	9.60	174.33	257.27	-82.94***
Shopping/getting service	13.35	23.65	-10.31***	11.66	31.12	-19.46***
Weaving/sewing/textile care	20.00	15.01	4.99	3.70	6.56	-2.85*
Cooking	74.25	72.50	1.75	81.63	88.25	-6.62
Domestic work	111.84	102.00	9.84	42.12	106.65	-64.53***
Care for children/adults/elderly	31.95	40.72	-8.76***	23.09	68.94	-45.84***
Travelling/commuting	43.76	31.25	12.51***	63.80	51.79	12.00**

Source: Authors' calculations based on WEAI 2.0 pilot data.
Notes: N=399 in Bangladesh; N=328 in Uganda. Paired *t*-tests performed comparing the mean time spent on each activity between methods. TD = time diary. SQ = stylized question. ***, **, and * denote statistical significance levels at 1%, 5%, and 10%, respectively.

Table 4. OLS regression results: Do differences in reported time use between time diary and stylized estimates vary according to the activity?

	Bangladesh	Uganda
Variable	Model 1A	Model 1B
Omitted category: Work as employed		
Own business work	17.766**	-12.909
	(8.501)	(8.490)
Farming/livestock/fishing	18.362*	-81.389***
	(10.129)	(12.404)
Shopping/getting service	-1.541	-17.907**
	(6.258)	(7.475)
Weaving/sewing/textile care	13.754*	-1.300
	(7.137)	(4.537)
Cooking	10.516*	-5.063
	(6.061)	(6.394)
Domestic work	18.609**	-62.973***
	(8.747)	(8.398)
Care for children/adults/elderly	0.000	-44.288***
	(6.655)	(7.365)
Travelling/commuting	21.278***	13.556*
	(6.899)	(7.074)
Constant	-8.765	-1.555
	(5.846)	(4.253)
Adjusted R-squared	0.006	0.065
Test of joint significance	8.217	24.336
Observations	3,591	2,952

Source: Authors' calculations based on WEAI 2.0 pilot data.
Notes: Dependent variable is the difference between time diary and stylized estimates. Household-level cluster-robust standard errors in parentheses. ***, **, and * denote statistical significance levels at 1%, 5%, and 10%, respectively.

Table 5. OLS regression results: Do differences in reported time use between time diary and stylized estimates vary according to respondent, enumerator, or interview characteristics?

	Bangladesh			Uganda		
Variable	Model 2A	Model 3A	Model 4A	Model 2B	Model 3B	Model 4B
Female	-0.264	2.605	2.620	-11.472***	-12.531***	-12.694***
	(2.918)	(4.197)	(4.188)	(4.181)	(4.181)	(4.246)
Age	0.346	0.465	0.462	-0.623	-0.613	-0.601
	(0.636)	(0.677)	(0.676)	(0.534)	(0.536)	(0.537)
Age-squared	-0.001	-0.003	-0.003	0.009*	0.010*	0.009*
	(0.006)	(0.007)	(0.007)	(0.005)	(0.005)	(0.005)
Literate	-8.567	-8.109	-8.203	-5.237	-4.579	-4.500
	(6.073)	(6.205)	(6.254)	(5.419)	(5.185)	(5.180)
Completed less than primary education	6.428	3.885	3.963	8.713	10.419*	10.401*
	(7.433)	(7.507)	(7.516)	(6.431)	(6.130)	(6.139)
Completed secondary education or above	2.384	-0.012	0.129	15.518**	13.341**	13.181**
	(6.733)	(7.168)	(7.234)	(6.240)	(6.162)	(6.185)
Interviewed alone	–	–	-1.270	–	–	-3.950
			(3.199)			(11.769)
Constant	-17.975	-4.086	-2.838	9.068	20.673	24.315
	(17.956)	(15.625)	(15.172)	(13.296)	(16.260)	(19.866)
Activity effects?	Yes	Yes	Yes	Yes	Yes	Yes
Enumerator effects?	No	Yes	Yes	No	Yes	Yes
Adjusted R-squared	0.007	0.011	0.010	0.071	0.077	0.077
Tests of joint significance:						
Activity effects	8.203***	8.150***	8.148***	24.286***	24.162***	24.154***
Respondent effects	1.859*	2.029*	2.024*	6.694***	6.624***	6.543***
Enumerator effects	–	57.850***	25.977***	–	2.903***	2.901***
Observations	3,591	3,591	3,591	2,952	2,952	2,952

Source: Authors' calculations based on WEAI 2.0 pilot data.
Notes: Dependent variable is the difference between time diary and stylized estimates. Household-level cluster-robust standard errors in parentheses. ***, **, and * denote statistical significance levels at 1%, 5%, and 10%, respectively.

Figure 1. Stylized time use questions from the Impact Evaluation of the Strengthening the Dairy Value Chain Project (SDVCP) in Bangladesh

Z6. Participation of household members in livestock activities and household activities

Livestock activity	Code	Who is primarily responsible to do it Member ID	Who does it in absence of the responsible person Member ID	In the past 30 days, who has usually done it? Member ID	In an average week in the past 30 days, how many hours total has this person spent on the activity?
Cleaning/draining of animal shed	1				
Taking care of Cows/animals					
Washing of animals	3				
Collecting/carrying fodder from field	4				
Taking animals to the field for grazing	5				
Purchasing feed	6				
Preparing feed and feeding the animals	7				
Taking the animals to hospital/somewhere else for treatment	8				
Taking the animals for vaccination and/or deworming	17				
Calling doctor for treatment of animals	9				
Taking the animals to hospital/somewhere for artificial insemination	10				
Calling doctor for artificial insemination of animals	11				
Milking animals	18				
Cleaning of milking area	12				
Cleaning and drying of utensils before and after milking	13				
Carrying the milk for selling	14				
Collecting the money after selling milk	15				
Grazing animals in the field	16				

Last four rows for purchase livestock and poultry are deleted as these are already asked in other modules.

Household activities	Code	Who is primarily responsible to do it Member ID	Who does it in absence of the responsible person Member ID	In the past 30 days, who has usually done it? Member ID	In an average week in the past 30 days, how many hours total has this person spent on the activity?
Feeding young children	1				
Looking after young children	2				
Cooking	3				
Washing clothes	4				

Source: Impact Evaluation of the Strengthening the Dairy Value Chain Project in Bangladesh, Endline Survey Questionnaire.

Figure 2. 24-hour recall time diary from the WEAI 2.0 Pilot in Bangladesh and Uganda

MODULE G6: TIME ALLOCATION

G6.01a: PLEASE RECORD A LOG OF THE ACTIVITIES FOR THE INDIVIDUAL IN THE LAST COMPLETE 24 HOURS (STARTING YESTERDAY MORNING AT 4 AM, FINISHING 3:59 AM OF THE CURRENT DAY). THE TIME INTERVALS ARE MARKED IN 15 MIN INTERVALS AND ONE TO TWO ACTIVITIES CAN BE MARKED FOR EACH TIME PERIOD BY DRAWING A LINE THROUGH THAT ACTIVITY. IF TWO ACTIVITIES ARE MARKED, THEY SHOULD BE DISTINGUISHED WITH A 1 FOR THE PRIMARY ACTIVITY AND 2 FOR THE SECONDARY ACTIVITY WRITTEN NEXT TO THE LINES. PLEASE ADMINISTER USING THE PROTOCOL IN THE INTERVIEWER MANUAL.

"Now I'd like to ask you about how you spent your time during the past 24 hours. This will be a detailed accounting. We'll begin from yesterday morning at 4am, and continue through to 4am of this morning."

		Night								Morning								Day																															
	Activity	4				5				6				7				8				9				10				11				12				13				14				15			
A	Sleeping and resting																																																	
B	Eating and drinking																																																	
C	Personal care																																																	
D	School (also homework)																																																	
E	Work as employed																																																	
F	Own business work																																																	
G	Farming/livestock/fishing																																																
J	Shopping/getting service (incl health services)																																																
K	Weaving, sewing, textile care																																																
L	Cooking																																																
M	Domestic work (incl fetching wood and water)																																																
N	Care for children/adults/elderly																																																
P	Travelling and communiting																																																
Q	Watching TV/listening to radio/reading																																																
T	Exercising																																																
U	Social activities and hobbies																																																
W	Religious activities																																																
X	Other, specify…																																																

Source: WEAI 2.0 Pilot individual questionnaire.

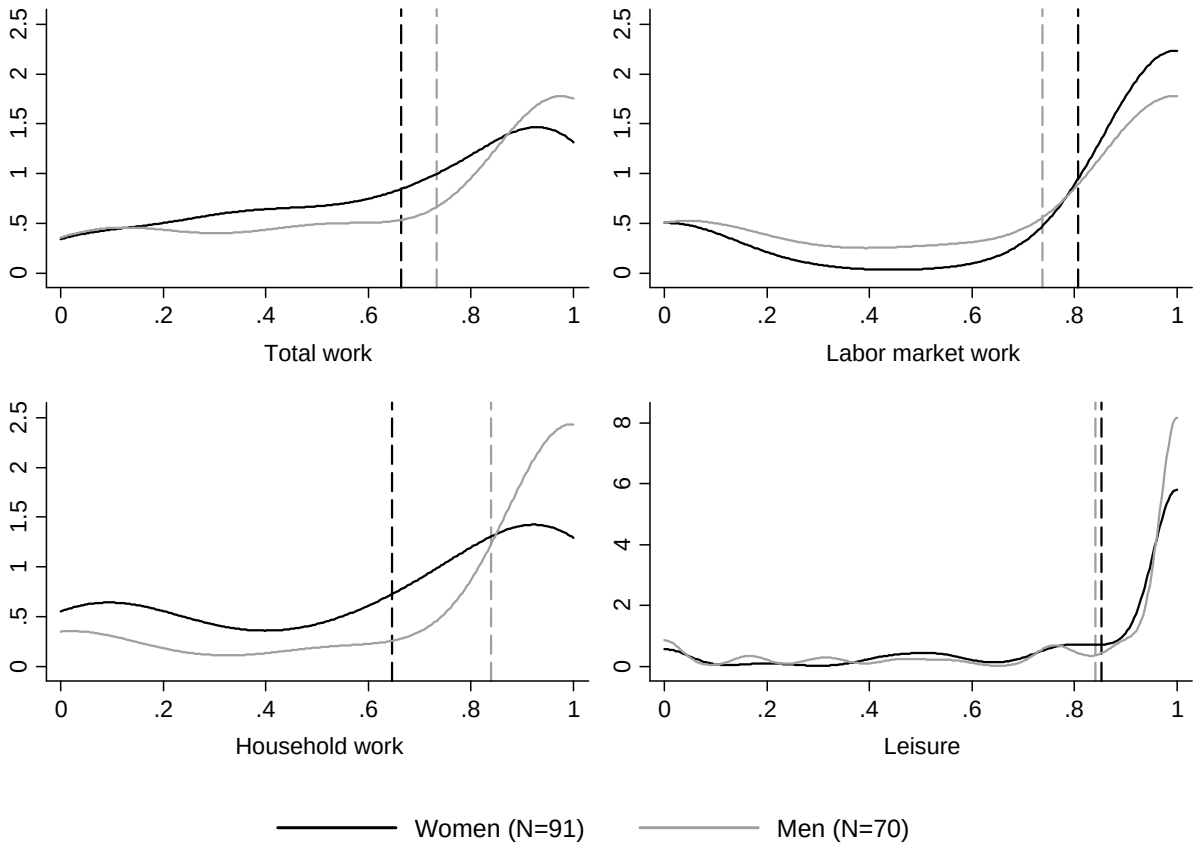
Figure 3. Stylized time use questions from the WEAI Pilot II in Bangladesh and Uganda

TIME CHECK: [__|__]:[__|__] CIRCLE ONE: AM/PM **MODULE G5: TIME ALLOCATION**

	G5.01: Now I'd like to ask you some questions regarding how you've spent your time over the last week and whether or not this was typical. **ENUMERATOR NOTE: THERE ARE 168 HOURS/WEEK.**	In the last 7 days, how much time in hours did you spend on [ACTIVITY]?	Did you spend a usual amount of time on [ACTIVITY] in the last 7 days?	Since the last week was not usual, within the last 6 months how much time do you usually spend on [ACTIVITY] per week?
		HOURS	**CIRLE ONE**	**HOURS**
A	Farming	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
B	Fetching water	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
C	Cooking	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
D	Care for children/adults/elderly	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
E	Work as employed (incl casual and salaried labor)	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
F	Own business work	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
G	Livestock/fishing	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
H	Market/getting service (including health services)	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
I	Weaving/sewing/textile care	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
J	Other domestic work	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
K	Fetching wood	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐
L	Traveling and commuting	☐☐☐	Yes.....1→ *NEXT ACTIVITY* No.....2 N/A.....98	☐☐☐

Source: WEAI 2.0 Pilot individual questionnaire.

Figure 4. Kernel density estimates of the proportion of time experienced as pleasant by gender and type of activity



Source: Figure 4.2 from Seymour and Floro (2016).
Notes: The dotted lines show the mean values among men and women, respectively. Only the difference for household work is statistically significant by conventional thresholds. Total work includes: work as employed; own business work; farming; construction; fishing; other work; collecting water; collecting firewood; vegetable gardening; animal husbandry; caring for children; caring for the sick/elderly; cooking; shopping/going to the market; cleaning the home; weaving, sewing, and textile care; and other domestic work. Labor market work includes: work as employed; own business work; farming; construction; fishing; and other work. Household work includes: collecting water; collecting firewood; vegetable gardening; animal husbandry; caring for children; caring for the sick/elderly; cooking; shopping/going to the market; cleaning the home; weaving, sewing, and textile care; and other domestic work. Leisure includes: traveling; watching television; listening to radio; reading; sitting with family; social activities; and other leisure.

Appendix

A. PROPOSED SURVEY EXPERIMENTS[19]

This section describes a series of survey experiments designed to assess the relative accuracy, cost effectiveness and scale-up feasibility of traditional and modern approaches to intra-household, individual-level data collection on time use in development settings.

To be able to rigorously evaluate the accuracy of different methods, we must be able to establish how respondents actually spend their time. We propose to do this via a self-administered, intensively supervised time diary, implemented over a two-day period, during which field staff will pay a minimum of three visits per day to the selected households to ensure that the respondents area complying with the survey protocols. The first day of implementation will be treated as trial, while the second day of data collection will inform our analyses. Depending on the level of literacy and familiarity with "clock" time among respondents, the self-reported diary may be pictorial in nature following Masuda et al. (2014). In our proposed experiments, this treatment would serve as our control group.

Activity-tracking devices

Several potential benefits might be gained by augmenting traditional time dairies with contemporaneous activity-tracking, for example activity-tracking data could be used to: (1) validate time diary accounts in post-fieldwork analysis; (2) inform how enumerators conduct time diary interviews (by providing bounds and checks on respondents' recall); and (3) gain insights into health effects of different activities or combinations of activities (i.e., how much energy is expended).

The experiment consists of combining a 24-hour recall time diary with sensor-based physical activity tracking monitoring. This treatment could be compared to both the control group and another treatment arm in which only a time diary is implemented. Each household will be visited three times by the enumerator over a period of five days. In Visit 1, enumerators will administer the core household and

[19] These designs draw on discussions with Talip Kilic from the World Bank.

individual questionnaires, sans time diary, and will supply each respondent with a wrist-worn, activity-tracking device. During this initial visit, any questions or concerns respondents may have about the devices will be answered to acclimatize the respondents to wearing the device. The respondents will be asked to wear the activity tracking device continuously for the next five days. Visit 2 will take place two days after Visit 1 to check compliance with wearing the activity tracker and to partially download the data stored on the activity trackers.

Visit 3 will take place two days after Visit 2 to do a second download of the data stored on the activity trackers. Depending on how fast data can be downloaded and made available to enumerators, data from Visits 2 or 3 might be used to assist enumerators (and respondents) during the time diary interview. Not only the information will be helpful in contextualizing the previous day (in particular towards possible increased identification of overlapping activities), it will also bind the recall period in a precise fashion, relying first and foremost on the sleep time data. At the conclusion of Visit 3, respondents will be asked a series of question about the extent to which they altered (or did not alter) their activities due to wearing activity tracking device.

"Priming" to improve recall

This experiment tests whether knowledge of an impending time diary interview can improve respondents' ability to accurately recall their actions. It will consist of two treatment arms, in addition to the control group. During Visit 1, respondents in Arm 1 will be notified that sometime over the next three days an enumerator will return to conduct a 24-recall time diary interview; additional or alternative strategies based on behavioral economics could be implemented at this stage. Visit 2 will take place on a randomly selected day among the next three days. During Visit 2, the enumerator will conduct a time diary interview, which will include at its conclusion a series of questions designed to assess whether the respondent changes his/her behavior based on the pre-interview warning. Time diary interviews will be conducted in Arm 2 households following standard protocols; respondents will not receive prior notification that a 24-recall time diary will be included in the interview.

SMS-based mobile phone surveys

One of the primary difficulties of using time diaries to collect time use data is that the data provided may not be representative of how they typically spend their time. That is, it is difficult to ascertain based solely on information from respondents themselves whether the previous day was typical or atypical.

This experiment investigates whether SMS-based mobile phone surveys might be used to cost-effectively collect time use data that accurately represents how a person spends their time. The experiment consists of asking respondents to complete short (five minutes or less) surveys through SMS text messages sent to their personal mobile devices at random moments over the course of a few weeks. Although the survey could be designed to collect information about several different aspects of a person's time use (e.g., location, persons with, emotions experienced), at a minimum it will ask respondents to describe how they spent the last hour of their time (in 15 minute intervals) and be designed to ensure that the person responding to the survey is the intended respondent (i.e., ensure that the phone has not changed hands).

The precise duration of the survey period (over how many days the SMS messages will be sent) and the frequency of the surveys will be designed based on cost and the researchers' determination of the number of data points necessary to construct a representative sample of the respondent's time. The entire experiment will ideally be staged so that it falls completely within a single agricultural season and preferably during a short enough period that similar activities are being performed to ensure that the average values that emerge from the exercise are meaningful. Mobile credit could be provided to respondents in exchange for timely compliance with the surveys.

References

Alkire, Sabina, Ruth Meinzen-Dick, Amber Peterman, Agnes R. Quisumbing, Greg Seymour, and Ana Vaz. 2013. "The Women's Empowerment in Agriculture Index." *World Development* 52: 71–91.

Arthi, Vellore, Kathleen Beegle, Joachim De Weerdt, and Amparo Palacios-López. 2016. "Not Your Average: Job Measuring Farm Labor in Tanzania." Policy Research Working Paper 7773. World Bank, Washington, DC.

Ayuso-Mateos, José Luis, Marta Miret, Francisco Félix Caballero, Beatriz Olaya, Josep Maria Haro, Paul Kowal, and Somnath Chatterji. 2013. "Multi-Country Evaluation of Affective Experience: Validation of an Abbreviated Version of the Day Reconstruction Method in Seven Countries." *PloS ONE* 8 (4). doi:10.1371/journal.pone.0061534.

Belli, Robert F. 1998. "The Structure of Autobiographical Memory and Event History Caldendar: Potential Improvement in the Quality of Retrospective Reports in Surveys." *Memory* 6 (4): 383–406.

Brown, Norman R. 2002. "Encoding, Representing, and Estimating Event Frequencies: A Multiple Strategy Perspective." In *Etc. Frequency Processing and Cognition.*, edited by Peter Sedlmeier and Tilmann Betsch, 37–53. Oxford; New York: Oxford University Press.

Brown, Norman R., Rebecca L. Williams, Erin T. Barker, and Nancy L. Galambos. 2007. "Estimating Frequencies of Emotions and Actions: A Web-Based Diary Study." *Applied Cognitive Psychology* 21: 259–76.

FAO. 2004. *Human Energy Requirements*. Rome: FAO. doi:92 5 105212 3.

Fisher, Kimberly. 2015. "Metadata of Time Use Studies." http://www.timeuse.org/information/studies/.

Floro, Maria S., and Anant Pichetpongsa. 2010. "Gender, Work Intensity, and Well-Being of Thai Home-Based Workers." *Feminist Economics* 16 (3): 5–44.

Frazis, Harley, and Jay Stewart. 2012. "How to Think about Time-Use Data: What Inferences Can We Make about Long- and Short-Run Time Use from Time Diaries?" *Annals of Economics and Statistics* 105/106: 231–45.

Hektner, J.M., J.A. Schmidt, and M. Csikszentmihalyi. 2007. *Experience Sampling Method: Measuring the Quality of Everyday Life*. Thousand Oaks, CA: SAGE Publications.

Juster, F. Thomas, Hiromi Ono, and Frank P. Stafford. 2003. "An Assessment of Alternative Measures of Time Use." *Sociological Methodology* 33: 19–54.

Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. 2004. "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method." *Science* 306 (5702): 1776–80. doi:10.1126/science.1103572.

Kan, Man Yee, and Stephen Pudney. 2008. "Measurement Error in Stylized and Diary Data on Time Use." *Sociological Methodology* 38: 101–32.

Krueger, Alan B., Daniel Kahneman, David Schkade, Norbert Schwarz, and Arthur A. Stone. 2009. "National Time Accounting: The Currency of Life." In *Measuring the Subjective Well-Being of Nations: National Accounts of Time Use and Well-Being*, edited by Alan B. Krueger, 9–86. Chicago, IL: University of Chicago Press.

Larson, Reed, and Mihaly Csikszentmihalyi. 1983. "The Experience Sampling Method." *New Directions for Methodology of Social and Behavioral Science* 15: 14–56.

Lawson, David. 2007. "A Gendered Analysis of 'Time Poverty'–The Importance of Infrastructure." GPRG Working Paper Series 078. http://economics.ouls.ox.ac.uk/12984/.

Loftus, Elizabeth F., and Wesley Marburger. 1983. "Since the Eruption of Mt. St. Helens, Has Anyone Beaten You up? Improving the Accuracy of Retrospective Reports with Landmark Events." *Memory & Cognition* 11 (2): 114–20. doi:10.3758/BF03213465.

Masuda, Yuta J., Lea Fortmann, Mary Kay Gugerty, Marla Smith-Nilson, and Joseph Cook. 2014. "Pictorial Approaches for Measuring Time Use in Rural Ethiopia." *Social Indicators Research* 115 (1): 467–82. doi:10.1007/s11205-012-9995-x.

Menon, Geeta. 1993. "The Effects of Accessibility of Information in Memory on Judgments of Behavioral Frequencies." *Journal of Consumer Research* 20: 431–40.

Miret, Marta, Francisco Félix Caballero, Arvind Mathur, Nirmala Naidoo, Paul Kowal, José Luis Ayuso-

Mateos, and Somnath Chatterji. 2012. “Validation of a Measure of Subjective Well-Being: An Abbreviated Version of the Day Reconstruction Method.” *PloS One* 7 (8): e43887. doi:10.1371/journal.pone.0043887.

Pontzer, Herman, David A. Raichlen, Brian M. Wood, Audax Z.P. Mabulla, Susan B. Racette, and Frank W. Marlowe. 2012. “Hunter-Gatherer Energetics and Human Obesity.” *PLoS ONE* 7 (7): 1–8. doi:10.1371/journal.pone.0040503.

Seymour, Greg, and Maria S. Floro. 2016. “Identity, Household Work, and Subjective Well-Being among Rural Women in Bangladesh.” IFPRI Discussion Paper 1580. International Food Policy Research Institute, Washington, DC.

Sudman, Seymour, and Norman M. Bradburn. 1973. “Effects of Time and Memory Factors on Response in Surveys.” *Journal of the American Statistical Association* 68 (344): 805–15.

UNDESA. 2010. “The World’s Women 2010: Trends and Statistics.” United Nations Department of Economic and Social Affairs, New York, NY.

van der Vaart, Wander, and Tina Glasner. 2007. “Applying a Timeline as a Recall Aid in a Telephone Survey: A Record Check Study.” *Applied Cognitive Psychology* 21: 227–38.

Williams, Jason R., Yuta J. Masuda, and Heather Tallis. 2015. “A Measure Whose Time Has Come: Formalizing Time Poverty.” *Social Indicators Research*. doi:10.1007/s11205-015-1029-z.