Article

Field Methods
2019, Vol. 31(1) 3-22

DOI: 10.1177/1525822X18797303
journals.sagepub.com/home/fmx

Times Have Changed: Using a Pictorial Smartphone App to Collect Time–Use Data in Rural Zambia

Thomas Daum[1], Hannes Buchwald[2], Ansgar Gerlicher[2], and Regina Birner[1]

Abstract

One challenge of collecting socioeconomic data, such as data on time-use, is recall biases. While time-use researchers have continuously developed new methods to make data collection more accurate and easy, these methods are difficult to use in developing countries, where study participants may have low literacy levels and no clock-based concepts of time. To contribute to the closing of this research gap, we developed a picture-based smartphone app called Time-Tracker that allows data recording in real time to avoid recall biases. We pilot tested the app in rural Zambia, collecting 2,790 data days. In this article, we compare the data recorded with the app to data collected with 24-hours recall questions. The results confirm the literature on recall biases, suggesting that using the app leads to valid results. We conclude that smartphone apps using visual tools provide new opportunities for researchers collecting socioeconomic data in developing countries.

[1] Hans-Ruthenberg-Institute of Agricultural Science in the Tropics, University of Hohenheim, Stuttgart, Germany
[2] Institute for Applied Science, University of Media, Stuttgart, Germany

Corresponding Author:
Thomas Daum, Hans-Ruthenberg-Institute of Agricultural Science in the Tropics, University of Hohenheim, Wollgrasweg 43, 70599 Stuttgart, Germany.
Email: thomas.daum@uni-hohenheim.de

Introduction

One major challenge of collecting data through surveys is recall biases, and there is strong evidence that recall biases can be substantial with regard to time-use (Chatzitheochari et al. 2017; Juster and Stafford 1991; Juster et al. 2003; Kan and Pudney 2008). In developing countries, time–use research is particularly challenging because study participants may have high rates of illiteracy and populations may lack clock-based concepts of time (Harvey and Taylor 2000; Kes and Hema 2006). In light of these challenges, time–use research in developing countries has often been neglected, which makes it difficult for governments and development practitioners to prioritize and design development programs and policies and to measure their effects. In contrast, time–use researchers working in developed countries have long focused on developing methods to collect more accurate time–use data. Recognizing that recall periods beyond two days lead to unreliable data, time–use researchers have considered daily time–use diaries to be the gold standard (Chatzitheochari et al. 2017; Juster et al. 2003). However, such time–use diaries are burdensome to fill and the administrative and processing costs are high (Chatzitheochari et al. 2017). This has led to calls for better time–use methods, potentially using electronic devices (Chatzitheochari et al. 2017; Minnen et al. 2014; Paolisso and Hames 2010; Seymour et al. 2017).

Following these calls, several research groups have developed app-based time–use diaries (Chatzitheochari et al. 2017; Fernee and Sonck 2014; Runyan et al. 2013). While these efforts have shown the potential of using smartphone apps for time–use research, these text-based apps are not applicable for research in developing countries. One noteworthy exception that aims to address the challenges of time–use researchers in developing countries is Masuda et al. (2014) who test piloted a pictorial diary set in Ethiopia. This set contains a book with a grid, activity stickers, and a clock, which beeps every 30 minutes. When the clock beeps, participants place a sticker in the book that reflects their then current activity. While being accessible for people with low or no literacy and without clock-based concepts of time, this method still seems cumbersome, and it does not allow for capturing simultaneous activities, which may, given the of 30-minute interval, consequently provide inaccurate data.

To sum up, time–use researchers in developing countries lack suitable data collection methods and, therefore, reliable data. To address the lack of suitable methods, we developed a picture-based smartphone application called Time-Tracker that allows study participants to record time-use in

Table 1. Advantages and Disadvantages of Methods to Collect Time–use Data.

Criteria	Surveys (Seasonal)	Surveys (Weekly)	Diaries (Daily/Weekly)	Observations (Real Time)	Apps (Real Time)
Recall bias	High	Medium/high	Low	No	Low
Social desirability bias	Yes	Yes	Yes	Yes	Low
Costs	Low	Medium/high[a]	Low	High	Medium
Respondent burden	Low	Medium	High	Medium/high	Low/medium

[a]Depending on whether questions are asked by phone or face-to-face (Arthi et al. 2018).

real time to avoid recall biases. The app can be combined with pop-up questions, a feature that we used to ask questions on quantity and quality of food consumed.

We used the app in Eastern Province of Zambia to collect approximately 2,790 data days on the time-use of farm families throughout an entire farming season. This was done as part of a larger study assessing the impacts of agricultural mechanization on intrahousehold labor allocation. This study was done because agricultural mechanization has received renewed attention in many developing countries (Daum and Birner 2017). However, the effects of mechanization on labor are ambiguous. The mechanization of activities that tend to be done by men, such as land preparation, may lead to land expansion and thereby a higher labor burden for nonmechanized activities that tend to be done by women, such as weeding.

We compared the data collected with the app, which we used as a benchmark, with answers from 24-hour recall questions. This comparison allowed us to explore how and why recall biases differ for different activities and for different household members. In brief, this article aims to contribute to the development of much needed methods to collect more accurate time–use data in developing countries.

Methodological Considerations

Data collection methods time–use researchers can adopt, including their advantages and disadvantages, are summarized in Table 1. They are further discussed in subsequent sections with a specific focus on their suitability for developing countries. Table 1 also depicts the *expected* advantages and disadvantages of using smartphone apps for collecting time–use data, such as the one introduced in this article.

Weekly and Seasonal Surveys

Most time–use studies rely on household surveys using recall questions such as: How much time did you spend last week/last farming season doing X? Using household surveys is inexpensive and allows for large sample sizes. However, the answers to survey questions "typically prove wide off the mark" (Juster and Stafford 1991: 482). Several aspects contribute to this, some of which are general problems and some of which are more pronounced in developing countries.

In general, study participants overestimate activities that are socially desirable and underestimate activities that are nondesirable and activities that they or the society do/does not perceive as work (Hofferth 1999; Juster and Stafford 1991; Juster et al. 2003). This is one reason why the length of activities may be reported differently by men and women. For example, Bianchi et al. (2012) found that men overestimate their contribution to household work by 70% in the United States. Study participants frequently overestimate secondary activities such as childcare (Juster et al. 2003). There is no clear agreement about whether sporadic activities are underestimated (Juster et al. 2003) or not (Menon 1993). The role of the intensity of different activities has not been studied much but may play role as well (Jodha 1988). There is a consensus that regular and externally structured activities, such as office work, have low biases (Juster et al. 2003).

Some of the challenges mentioned are more severe in developing countries, with regard to agriculture, for the following reasons. First, study participants may lack a clock-based concept of time. Second, compared to people from developed countries, people from developing countries tend to have less-structured days, which makes recalling time-use more difficult (Arthi et al. 2018). Third, the seasonality of farming may have effects on the perception of time spent on activities that are performed highly irregularly (Arthi et al. 2018). In view of these challenges, it is problematic that rural livelihood surveys frequently use postharvest recall questions that cover the entire farming season.

Arthi et al. (2018) found that Tanzanian farmers report a work time that is four times higher when asked via a postharvest instead of a weekly survey, which suggests that the long-standing debate on whether small or large farms are more efficient and thus whether agricultural development should be based on small or large farms (Collier and Dercon 2014; Larson et al. 2014) may be based on unreliable data. One reason for the overestimation of farm labor may be that postharvest recall questions force study participants to make "cognitively taxing calculations which result in labour inferences that appear to be based on recent rather than representative experiences" (Arthi et al. 2018:19). They also speculate that, in view of the harvest produced, labor can be overstated during good harvests and understated during bad harvests. Another problem is that agricultural surveys are often designed to be answered by the "household head" who may underestimate the work contribution of his or her kin.

Using weekly recall questions does substantially reduce the recall biases associated with household surveys (Arthi et al. 2018). However, they still do not meet the implicit standard of time–use researchers who argue that recall period beyond two days leads to unreliable data. Also, weekly data collection may be associated with high costs unless study participants are contacted by phone, which may lead to excluding study participants without phones.

Usually, comparing recall data is difficult because not only are different types of activities recalled as having lasted different times they are also recalled differently by different genders, social groups, and people from different countries with different familiarities with clock-based concepts of time. This makes intrahousehold comparisons of time-use very difficult. It also makes comparisons between different social groups challenging, for example, farmers and pastoralists, or Germans and Ghanaians.

Time–Use Diaries

Studies may also use time–use diaries in which study participants fill out 24-hour time grids that are divided into 15- or 30-minute slots either freely or using precoded activities. Time–use diaries are considered the most reliable and accurate data collection method, as they are less prone to recall problems as well as social desirability bias (Chatzitheochari et al. 2017; Juster et al. 2003; Paolisso and Hames 2010). It has been argued that time–use diaries are "the only valid measurement of time-use, and less expensive substitutes are substantially lower quality and have systematic biases" (Juster and Stafford 1991:482).

However, time–use diaries involve text-based questions and are burdensome to complete (Chatzitheochari et al. 2017). Therefore, diaries are not a viable option for developing countries unless they are filled with the help of interviewers, which may lead to biases. An exception is the above-described pictorial time diaries used by Masuda et al. (2014). While the use of pictures allows low-literacy and illiterate participants to use these diaries, they are still cumbersome to fill. In addition, they are based on 30-minute slots, which may affect study participants to underreport activities that are regularly performed throughout the day but are shorter than 30 minutes each time they are performed (Chatzitheochari et al. 2017; Kelly et al. 2015).

Direct Observations

Direct observations can eliminate recall biases and address the illiteracy problem (Kes and Hema 2006; Paolisso and Hames 2010); however, direct observations are expensive and thus reduce potential sample sizes (Harvey and Taylor 2000; Kes and Hema 2006). Assuming that a trained research assistant costs US$30/day (working maximum of eight hours/day), the cost of observing 2,790 days (the number of days recorded in this study) would have been US$167,000 (not including the costs for organization, cars, and accommodation). In addition, the observer's presence may affect the behavior of the observed, the so-called Hawthorne Effect (Kes and Hema 2006; Paolisso and Hames 2010).

Method

Incorporating strengths of some of the abovementioned time–use methods while overcoming some of their challenges, we developed an open-source smartphone app called Time-Tracker (the code can be accessed through https://github.com/HannesBuchwald/TimeTracker), which allows study participants to record data themselves. To ensure that study participants can easily record and capture the social context of the study area, the app was developed in close collaboration with farmers. To guarantee that populations with low or no literacy can participate, the app works only with pictures; data recording was designed to be as simple as possible to allow study participants with no experience with mobile phones or smartphones to effectively and easily use the app and to make sure that study participants do not develop "entry fatigue," losing the motivation to carefully record data. Figure 1 shows the main screen of the app, which displays pictures of 88 typical daily activities that we selected and designed together with the local population.

To start recording an activity, study participants click on the respective picture (e.g., hand hoeing): up to three activities can be recorded at the same time to capture simultaneity (e.g., hand hoeing and caring for a baby). To stop recording an activity, the participant clicks on the respective picture again. Activities are thus recorded in real time, which reduces recall biases.

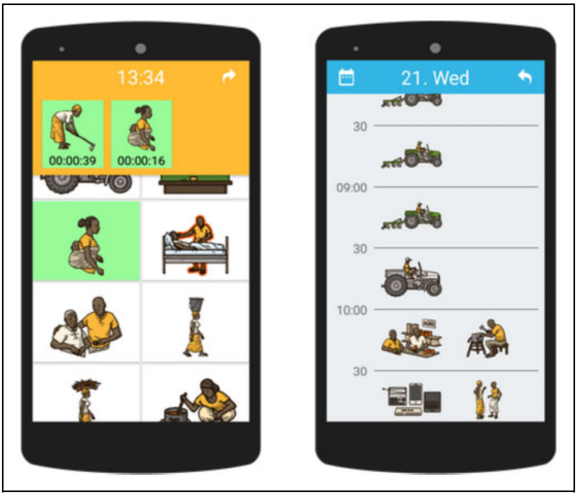

Figure 1. The data entry screen (left). The screen on the right allows study participants to see the recorded data. The researchers can activate a hidden button and correct potential mistakes.

Activities participants are performing at that moment are displayed on top of the screen.

The study participants can also indicate whether the work was done on their own field or on the field of others as agricultural laborers (with a triple click on the respective drawing). However, some study participants had difficulties with this mechanism. The research team, therefore, frequently cross-checked with the participants whether work was done on the field of others.

In addition to time–use data, study participants can record food intake. To make this possible, we designed a plug-in that opens when the activity "eating and drinking" was terminated. Study participants are then shown four differently filled plates to record the quantity of food consumed. Afterward, they are shown different food groups (e.g., cereals, roots, and vegetables), which allows them to indicate the diversity of the food they ate.

Before study participants used the app, we introduced them to it. First, we practiced the use of touch screens; then, we clarified questions about the pictures. Moreover, the study participants practiced using the app with the help of explanatory stories. For example, using their local language, they were told to record the following story using the app: "Christian goes to the field to hoe while listening to the radio; one the way back, he uses the bicycle of a friend. After reaching home, he eats vegetables and *nshima*" (a maize dish in Zambia).

The participants were given a smartphone with the app for the data-recording period. We lent them a smartphone to avoid selection biases, which would have occurred if we had only selected study participants who already owned a smartphone. We used 50 smartphones costing US$90 each, which allowed us to work with 45–48 respondents at a time. In total, the costs of developing and administering the app were around US$40,000 (including the smartphone costs). Thus, each collected datum cost US$13.4.

The smartphones were configured, so that only the time–use app could be used. Blocking other smartphone functions was done to ensure that smartphone use did not alter daily routines of the participants. The blocking had two positive side effects. First, it may have reduced any temptation to "lose" the smartphone. Second, the blocking enabled only the app to run, which extended the battery life up to five days. When the battery level was below 50%, we distributed power banks.

To ensure that the smartphones worked properly, the research team made daily random checks. This approach also allowed the team to double-check whether study participants recorded the activities that they actually performed and to help participants manually enter activities that they may have forgotten. These corrections were possible through a second screen of the smartphone app, which is shown in Figure 1. Data recording and submission were done off-line, and after the recording phase, the research team uploaded the data from the smartphone to a laptop using a local Wi-Fi network. The participants received small gifts such as caps and fabrics as appreciation for their participation in the study.

Study Site and Sampling

The study was conducted as part of a larger research project that aimed at assessing the effects of agricultural mechanization in Eastern Province of Zambia. For this, we had used a two-stage sampling procedure to select 62 farm families with different mechanization levels based on the population of the nationally representative Rural Agricultural Livelihood Survey. The households were located across four different communities, all of which have been dominated by smallholder farmers. On average, farmers cultivate 2.3 hectares—mainly maize but also cotton, sunflower, groundnuts, and tobacco. Farming is characterized by a short rainy season and an extensive dry season. Most of the farming activities are done manually (1% of the households use, own, or hired mechanical power for land preparation and

Table 2. Sample Characteristics.

Variable	Mean
Sample size	62
Household size (members)	7.1
Age male (years)	47.4
Age female (years)	39
Age app using child (years)	15.6
Education level	8.9
Land size cultivated (hectares)	5.2
Farm income	US$1,532

Note: Education levels range from 1 = first grade to 20 = master degree.

57% use animal traction). Land and labor productivity is low, and 90% of the rural population live on less than US$1.25/day (all above from Indaba Agricultural Policy Research Institute 2016). Table 2 shows selected sample characteristics. In each of the selected households, the head of the household, one spouse, and one child (alternating between boys and girls) were trained to use the smartphone app. After the training, data were collected at five different stages during the 2016–2017 season: land preparation, planting, weeding, harvesting, and processing. At each point, households recorded their time-use for three days.

This study focuses on the last two stages, harvesting and processing. During these stages, study participants were also asked 24-hour recall questions after the last day of using the smartphone app. Twenty-four-hour recall questions made it possible to compare data recorded with the app to data elicited from recall questions. The recall questions asked were about major time–use categories such as farming or household chores (giving clear examples of different time–use categories). To reduce potential "adding up" problems, the day to be recalled was split into five time–use intervals.

Results

All sampled household members were able to use the app, which indicates that the use of a well-designed and explained smartphone app for data collection does not lead to selection bias. Figure 2 shows the average age of study participants. The wide range of study participants (from seven to 92 years) suggests that all age-groups can use the Time-Tracker. Also, participants with low or no literacy were able to use the app as well as study participants who had no prior exposure to smartphones.

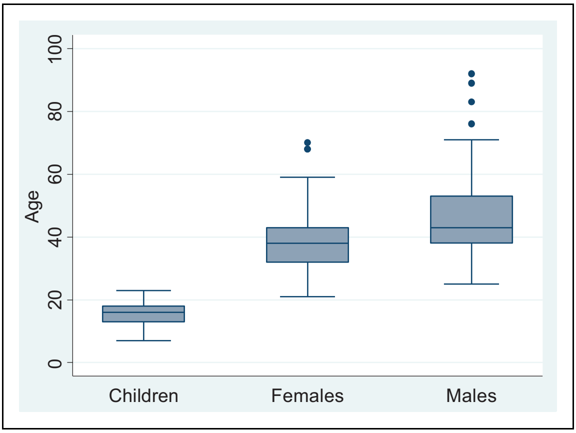

Figure 2. Age distribution of participants.

Table 3. Percentage of Data Entered/Changed by Enumerators.

Data Collection Rounds	Data Entered/Corrected by Research Team because Participants Forgot Entering or Clicked Wrong Activity (%)
Land preparation	0.6
Planting	0.8
Weeding	0.7
Harvesting	2.9
Processing	2.3

Our experiences suggest that study participants enjoyed working with the app, handled the smartphones with care, and took data recording very seriously. The data collected with the smartphone app appear to be of high quality and this can be seen from Table 3, which indicates that only a small percentage of the data had to be entered or corrected by the research team who supervised data collection.

Table 4 shows an example of a typical data day, showing both the time-use and food intake. During the entire data collection phase, only one smartphone disappeared and one accidently cracked, but still worked.

Table 5 compares durations of different time–use categories recorded by participants using the app and as they are recalled the next day. The differences indicate the magnitude of recall biases of 24-hour recall questions.

Table 4. Example of a Data Day.

						Food				
Activity	ID	Date	Start	End	Piecework	Quantity	Cereals	Vegetables	. . .	Diversity Score
Sleeping	1	May 06, 2017	00:00:00	05:34:28	No					
Personal Hygiene		(Saturday)	05:34:51	05:46:48	No					
Walking (unloaded)			05:47:07	06:02:07	No					
Harvesting (manual)			06:02:27	12:01:59	No					
Walking (unloaded)			12:02:08	12:28:22	No					
. . .			. . .	. . .	. . .					
Personal hygiene			19:34:30	19:47:53	No					
Eating + drinking			19:47:59	20:08:57	No	1	1	1	1	3
Chatting			20:09:46	20:31:38	No					
Sleeping			20:31:50	00:00:00	No					

Table 5 shows that some activities are recalled as lasting significantly longer (overestimated) compared to the time recorded with the smartphone app. For example, farming activities are recalled as having lasted between 29% and 47% longer. Of activities perceived as lasting significantly shorter (underestimated), social life activities, which were recalled as lasting 45–54% less than the app-recorded time, had the biggest difference.

Table 6 combines rounds I and II but splits the study participants into different categories (males, females, and children). It shows that activities such as farming, household chores, and social activities are gendered. Table 6 suggests that different respondent groups recall different activities with differing accuracies. For example, construction seems to be well recalled by females and children but recalled as lasting much longer than app-recorded time by males. Males did recall their contribution to household chores rather accurately, whereas, surprisingly, women did not.

We also analyzed the time–use difference, treating the difference between recall estimates and app-recorded data as a measurement error in recall data. These errors are normally distributed for all time–use categories. We tested whether the measurement error in recall data is significantly correlated with age and education but found no evidence (see Figure 3).

In the Online Supplemental Material, we indicate the accuracy of recall data both graphically and by reporting the mean squared standard deviation (MSD) around the 1:1 line for different demographic characteristics (male, female, and children). As suggested by Gauch et al. (2003), we separated the MSD into three categories to obtain a deeper understanding into recall accuracy: squared bias, showing translation; nonunity slope, showing rotation; and lack of correlation, showing scatter. We also graphically indicated whether the magnitude of errors vary systematically with the size of recall estimates. The results suggest that the longer an activity's duration was recalled, the more likely participants overestimated the activity's duration, which was the case for all time–use categories (not only the ones shown here). This observation seems more pronounced with regard to farming and household chores (a and b) than social life (c).

Discussion

Comparative Advantage over Existing Methods

The Time-Tracker combines advantages of existing time–use data collection methods and overcomes their respective disadvantages. The app allows study participants to record time-use in real time, which reduces the recall

Table 5. Comparison of Time-Use Recorded with App and 24-hour Recall Questions in Two Data Collection Rounds (Harvesting and Processing).

Minutes per activity (Daily)	Round I			Round II		
	Real Time (App)	24-hour Recall	Difference (%)	Real Time (App)	24-hour Recall	Difference (%)
Farming	200.3 (15.0)	258.0 (16.5)	29***	149.9 (14.9)	220.9 (17.9)	47***
Expanded farming	26.5 (6.8)	30.5 (9.4)	15	47.3 (9.3)	21.0 (6.3)	−56**
Off-farm work	11.2 (6.6)	18.2 (7.9)	63	11.5 (5.8)	12.4 (6.0)	8
Community work	2.2 (2.0)	3.0 (3.0)	36	4.2 (2.7)	3.7 (3.0)	−12
Meeting	7.5 (5.1)	19.8 (5.6)	164	4.2 (2.2)	8.3 (2.6)	98
Mobility	146.8 (12.6)	101.4 (9.9)	−31***	121.6 (11.4)	89.8 (7.2)	−26**
Education	25.9 (8.8)	28.4 (7.9)	10	36.0 (10.1)	34.5 (9.5)	−4
Care for others	24.2 (6.1)	118.5 (14.7)	390***	26.1 (7.8)	103.7 (12.7)	297***
Household chores	125.8 (12.4)	95.8 (9.8)	−24*	124.5 (11.3)	107.2 (10.5)	−14
Construction	5.5 (3.8)	9.1 (3.3)	65	6.1 (3.9)	16.9 (5.7)	177
Sleeping	710.3 (11.2)	608.0 (11.5)	−14***	718.5 (11.7)	603.7 (13.7)	−16***
Personal care	77.7 (4.3)	62.3 (3.4)	−20***	75.1 (4.4)	65.9 (5.2)	−12
Social life	172.2 (15.7)	77.9 (9.2)	−55***	193.3 (16.2)	105.2 (9.2)	−46***

Note: The values of standard deviations are given in brackets.
Asterisks indicate significantly differing mean values.
***$p < .01$.
**$p < .05$.
*$p < .1$.

Table 6. Comparison of Time-Use Recorded with App and 24-hour Recall Questions by Different Respondents.

	Round I and Round II								
	Males			Females			Children		
Minutes per Activity (Daily)	Real Time (App)	24-hour Recall	Difference (%)	Real Time (App)	24-hour Recall	Difference (%)	Real Time (App)	24-hour Recall	Difference (%)
Farming	184.2 (19.5)	247.7 (21.4)	34**	213.1 (17.6)	270 (20.4)	27**	117.4 (17.1)	191.8 (20.8)	63***
Expanded farming	77.7 (13.5)	65.6 (14.9)	−16	12.6 (5.4)	2.7 (1.4)	−79*	20.8 (8.7)	9.3 (7.1)	−55
Off-farm work	21.9 (10.0)	34.3 (13.0)	57	4.9 (4.8)	10.3 (6.0)	110	7.4 (7.4)	0 (0)	−100
Community work	6.4 (4.0)	5.5 (4.4)	−14	2.7 (2.7)	4.2 (4.0)	56	0.1 (0.1)	0 (0)	−100
Meeting	14.9 (8.0)	22.0 (7.5)	48	1.3 (1.0)	9.2 (3.5)	608**	1.2 (1.1)	11.1 (4.3)	825**
Mobility	197.3 (18.3)	112.5 (10.4)	−43***	89.7 (10.5)	97.3 (11.7)	8	118.2 (12.1)	74.1 (8.5)	−37***
Education	0 (0)	0 (0)	0	2.8 (2.3)	3.6 (3.3)	29	101.6 (20.6)	102.2 (18.5)	1
Care for others	6.2 (3.3)	42.6 (8.5)	−587***	51.1 (11.3)	180.4 (18.2)	253***	14.2 (7.0)	102.2 (18.7)	620***
Household chores	9.9 (2.9)	12.1 (2.7)	22	233.3 (13.0)	172.2 (12.5)	−26***	120.8 (13.9)	114.4 (12.6)	−5
Construction	10.2 (6.6)	36.4 (9.3)	257***	1.9 (1.9)	1.1 (0.6)	−42	5.6 (4.8)	1.3 (0.9)	−77
Sleeping	732 (13.5)	657.8 (11.7)	−10***	702.1 (12.4)	604 (13.4)	−14***	709.8 (16.6)	549.7 (19.7)	−23***
Personal care	69.5 (3.9)	76 (7.5)	9	77.8 (3.8)	60.2 (3.5)	−23***	82.4 (8.0)	55.3 (4.1)	−33***
Social life	205.1 (22.6)	109.7 (12.0)	−47***	140.9 (14.1)	73.3 (9.1)	−48***	209.6 (21.5)	93.8 (13.1)	−55***
Sample size	99	99		109	109		87	87	

Note: The values of standard deviations are given in brackets.
Asterisks indicate significantly differing mean values.
***$p < .01$.
**$p < .05$.
*$p < .1$.

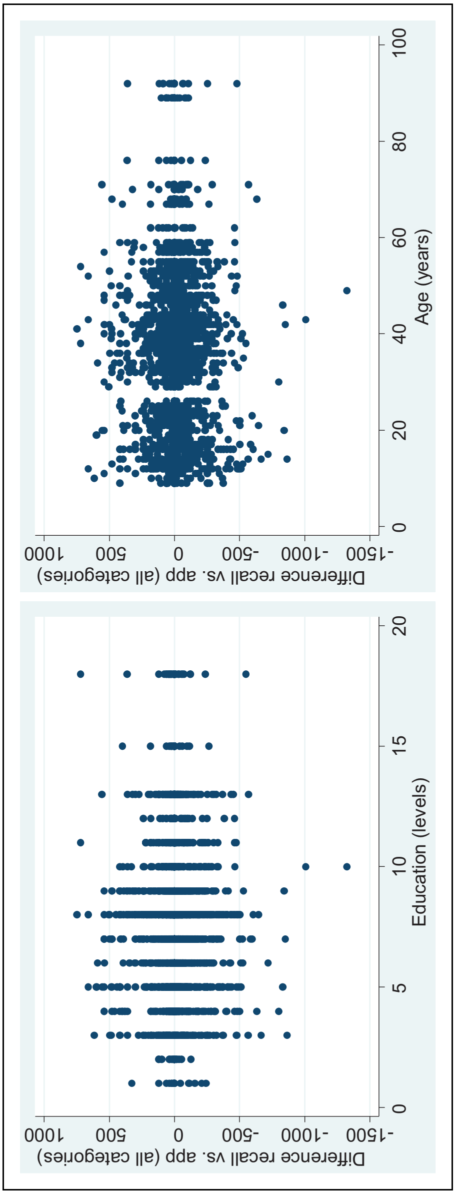

Figure 3. Recall error by educational level and age.

biases associated with household surveys. Using pictures, the app overcomes the text hurdle of existing paper and electronic time–use diaries. Also, the smartphone app has overcome some drawbacks of the pictorial diary sets of Masuda et al. (2014) described above: Instead of recording activities based on a 30-minute time slot, which inevitably introduces inaccuracies, especially for tasks that last less than 30 minutes, data recording with the smartphone app is straightforward and can be done in real time. We found that the study participants recorded data with great care and accuracy. We also found that the use of a smartphone app did not lead to selection bias, as all members of all sampled households were able to use the app. This suggests that well-designed smartphone apps may be used by time–use researchers to collect data in developing countries.

The Role of Recall Biases

To validate the data collected with the Time-Tracker, we compared the data recorded using the app with data collected through widely used 24-hour recall questions. We find that regular and uniformly structured activities such as going to school are recalled as lasting the same time as recorded with the smartphone app. This also validates that the data recorded with the app are accurate. Confirming existing literature, we find that the length of socially desirable (and arduous) activities, such as farming, is overstated. This confirms Arthi et al. (2018), who found that reported farm labor decreased by a factor of four when the recall period was reduced from postharvest data collection to weekly data collection. Our results (moving from 24-hour recall questions to real-time data) suggest that actual working time may even be lower. This implies that the labor productivity of farm households may be much higher than commonly estimated based on recall studies.

We found that, similar to farming time, the duration of care activities was reported to be longer than the duration from recall questions. This could partly be explained by findings in the literature, which suggest that durations of secondary activities, such as care taking, are often overestimated (Juster et al. 2003). However, study participants may not have recorded all care activities, especially when they lasted only short periods of time. Confirming Hofferth (1999) and Juster et al. (2003), we find that recalled durations of social life activities are significantly shorter than durations recorded with the app. The reason to understate the duration of such activities may be that study participants perceive these activities to be less desirable by society. Interestingly, male household members were not found to overestimate their contribution to household chores, a contrast with the findings of Bianchi et al. (2012) from the United States. Potentially, the difference is due to the fact that household chores are not seen as desirable for males in Zambia.

Limitations and Directions for Future Research

It is important to note that the time–use difference reported may be underestimated because the study participants, who had used the Time-Tracker the previous days, were already sensitized about time-use during recall questions. To avoid the inherent difficulty of comparing two data collection methods at the same time, one would need to conduct a randomized control trial as done by Arthi et al. (2018).

By using the Time-Tracker at various times of the farming season, we were able to capture seasonality. However, data recording only at specific (even if well-chosen) points of time does not make it possible to extrapolate the data over the entire farming season. For example, a participant may weed her fields 120 minutes/day for 15 days (a total of 1,800 minutes), whereas another may weed her fields 90 minutes/day for 30 days (a total of 2,700 minutes). Looking at specific data points (such as one day), one may wrongly conclude that the second participant works less. Data collection using a postharvest questionnaire would capture the difference in days weeded, but since it is based on recall questions, this information may be inaccurate. This implies a need for further research to find ways to collect data over extended periods of time while ensuring that study participants do not develop a fatigue. One solution may be gamification of the app, making app usage more attractive by using game-design elements, or allowing study participants to collect airtime credits. Furthermore, research can aim to find ways to better extrapolate data for the entire farming season.

There are additional ideas on how to further improve such smartphone applications. So far, similar to most time–use research, the Time-Tracker does not capture the intensity of efforts. This shortcoming could be addressed by combining the app with fitness trackers. Also, the Time-Tracker does not indicate whether activities are perceived as enjoyable; absence of this function can be improved by asking participants whether they enjoyed the activity with the help of pop-up windows (similar to Fernee and Sonck 2014). Thinking some steps ahead, data collected with smartphone apps may be validated using cameras and built-in sensors.

Recommendations

While we recommend using smartphone apps for time–use researchers in developing countries, several aspects need to be kept in mind. First, the app design as well as the selection and drawing of the illustrations need to be done in close collaboration with study participants. Second, when introducing the app, it is important to consider the role of village authorities, social dynamics, and beliefs. For example, it is key to explain how participating households are selected to avoid social tensions. Third, it is important to have sufficient training on how to handle smartphones and the app.

Conclusion

While time–use researchers have been continuously developing more accurate and user-friendly methods to collect time–use data, these efforts have been largely restricted to the developed world. This skewness has resulted in a lack of methods that could be adopted to collect time–use data in developing countries where study participants may have low literacy levels and no clock-based concept of time. In this article, we presented a picture-based smartphone app that allows researchers to collect time–use data with high precision in developing countries. The results suggest that well-tailored smartphone apps that use visual tools provide new and much needed pathways for time–use researchers working in developing countries.

Authors' Note

The study was approved by the ethics committee of the University of Hohenheim. All study participants gave their written consent to participate. For participants below the age of 18, consent was obtained from themselves and their guardian.

Acknowledgments

The authors are grateful to all farm families participating in the study. The authors are thankful for financial support from the Program of Accompanying Research for Agricultural Innovation, which is funded by the German Federal Ministry of Economic Cooperation and Development.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article is funded by The Program of Accompanying Research for Agricultural Innovation, which is funded by the German Federal Ministry of Economic Cooperation and Development.

Supplemental Material

Supplemental material for this article is available online.

References

Arthi, V., K. Beegle, J. de Weerdt, and A. Palacios-López. 2018. Not your average job: Measuring farm labor in Tanzania. *Journal of Development Economics* 130:160–72.

Bianchi, S. M., L. C. Sayer, M. A. Milkie, and J. P. Robinson. 2012. Housework: Who did, does or will do it, and how much does it matter? *Social Forces* 91: 55–63.

Chatzitheochari, S., K. Fisher, E. Gilbert, L. Calderwood, T. Huskinson, A. Cleary, and J. Gershuny. 2017. Using new technologies for time diary data collection: Instrument design and data quality findings from a mixed-mode pilot survey. *Social Indicators Research* 137:379–90.

Collier, P., and S. Dercon. 2014. African agriculture in 50 years: Smallholders in a rapidly changing world? *World Development* 63:92–101.

Daum, T., and R. Birner. 2017. The neglected governance challenges of agricultural mechanisation in Africa—Insights from Ghana. *Food Security* 9:959–79.

Fernee, H., and N. Sonck. 2014. Measuring smarter: Time-use data collected by smartphones. *Electronic International Journal of Time-Use Research* 11:94–96.

Gauch, H. G., J. T. Hwang, and G. W. Fick. 2003. Model evaluation by comparison of model-based predictions and measured values. *Agronomy Journal* 95: 1442–46.

Harvey, A. S., and T. Taylor. 2000. *Designing household survey questionnaires for developing countries: Lessons from 15 years of the Living Standards Measurement Survey*. Washington, DC: The World Bank.

Hofferth, S. 1999. *Family reading to young children: Social desirability and cultural biases in reporting. National Research Council Workshop on measurement and research on time-use*. Washington, DC: National Research Council.

Indaba Agricultural Policy Research Institute. 2016. *Rural Agricultural Livelihoods Survey 2015 Survey Report*. https://images.agri-profocus.nl/upload/post/RALS_2015_Survey_Report__Finalv-_edited1456827249.pdf (accessed January 8, 2018).

Jodha, N. S. 1988. Poverty debate in India: A minority view. *Economic and Political Weekly* 23:2421–28.

Juster, F. T., H. Ono, and F. P. Stafford. 2003. An assessment of alternative measures of time-use. *Sociological Methodology* 33:19–54.

Juster, F. T., and F. P. Stafford. 1991. The allocation of time: Empirical findings, behavioural models, and problems of measurement. *Journal of Economic Literature* 29:471–522.

Kan, M. Y., and S. Pudney. 2008. Measurement error in stylized and diary data on time-use. *Sociological Methodology* 38:101–32.

Kelly, P., E. Thomas, A. Doherty, T. Harms, Ó. Burke, J. Gershuny, and C. Foster. 2015. Developing a method to test the validity of 24 hour time-use diaries using wearable cameras: A feasibility pilot. *PLoS One* 10:e0142198.

Kes, A., and S. Hema. 2006. In *Gender and time poverty in sub-Saharan Africa*. World Bank Working Paper No. 73, Washington, DC, pp.13–38.

Larson, D. F., K. Otsuka, T. Matsumoto, and T. Kilic. 2014. Should African rural development strategies depend on smallholder farms? An exploration of the inverse-productivity hypothesis. *Agricultural Economics* 45:355–67.

Masuda, Y. J., L. Fortmann, M. K. Gugerty, M. Smith-Nilson, and J. Cook. 2014. Pictorial approaches for measuring time-use in rural Ethiopia. *Social Indicators Research* 115:467–82.

Menon, G. 1993. The effects of accessibility of information in memory on judgments of behavioural frequencies. *Journal of Consumer Research* 20:431–40.

Minnen, J., I. Glorieux, T. van Tienoven, S. Daniels, D. Weenas, J. Deyaert, S. Van den Bogaert, and S. Rymenants. 2014. Modular online time-use survey (MOTUS)—Translating an existing method in the 21st century. *Electronic International Journal of Time-Use Research* 11:73–93.

Paolisso, M., and R. Hames. 2010. Time diary versus instantaneous sampling: A comparison of two behavioral research methods. *Field Methods* 22:357–77.

Runyan, J. D., T. A. Steenbergh, C. Bainbridge, D. A. Daugherty, L. Oke, and B. N. Fry. 2013. A smartphone ecological momentary assessment/intervention "app" for collecting real-time data and promoting self-awareness. *PLoS One* 8:e71325.

Seymour, G., H. J. Malapit, and A. Quisumbing. 2017. *Measuring time-use in development settings*. Policy Research Working Paper No. 8147, World Bank, Washington, DC.