Development Engineering 5 (2020) 100056

Contents lists available at ScienceDirect

Development Engineering

journal homepage: http://www.elsevier.com/locate/deveng

Using sensors to measure technology adoption in the social sciences

Adina Rom [a,*], Isabel Günther [b], Yael Borofsky [b]

[a] ETH4D, ETH Zurich, Clausiusstrasse 37, 8092, Zurich, Switzerland
[b] Development Economics Group, ETH Zurich, Switzerland

ARTICLE INFO

Keywords:
Sensor
Self-report surveys
Measurement error
Social desirability bias
Technology adoption
Hawthorne effect

ABSTRACT

Empirical social sciences rely heavily on surveys to measure human behavior. Previous studies show that such data are prone to random errors and systematic biases caused by social desirability, recall challenges, and the Hawthorne effect. Moreover, collecting high frequency survey data is often impossible, which is important for outcomes that fluctuate. Innovation in sensor technology might address these challenges. In this study, we use sensors to describe solar light adoption in Kenya and analyze the extent to which survey data are limited by systematic and random error. Sensor data reveal that households used lights for about 4 h per day. Frequent surveyor visits for a random sub-sample increased light use in the short term, but had no long-term effects. Despite large measurement errors in survey data, self-reported use does not differ from sensor measurements on average and differences are not correlated with household characteristics. However, mean-reverting measurement error stands out: households that used the light a lot tend to underreport, while households that used it little tend to overreport use. Last, general usage questions provide more accurate information than asking about each hour of the day. Sensor data can serve as a benchmark to test survey questions and seem especially useful for small-sample analyses.

1. Introduction

Since the 1980s, advances in research design and analytical tools have increased the scientific impact and policy relevance of applied microeconomics, which Angrist and Pischke (2010) called a "credibility revolution." The increased use of natural experiments and randomized controlled trials (RCTs) were of particular importance to this development (Duflo et al., 2008; Angrist and Pischke, 2010). Alongside this trend, there has been an increase in the collection of household-level survey data. While methodological advances have been remarkable, much of the research in applied microeconomics in low-income countries still relies heavily on self-reported survey data, which are prone to measurement errors and can be expensive to collect. In pursuit of ways to mitigate measurement errors in survey data and thanks to recent technological breakthroughs, social scientists have started to turn to entirely new types of data, such as satellite imagery, cortisol stress tests, cell phone network data, and sensors, as a means of complementing self-reported survey data and, hopefully, improving the accuracy and precision of measurements. In this paper, we analyze how sensor data compares to household survey data on technology adoption, in this case, solar light usage in rural Kenya.

Some of the most discussed challenges associated with self-reported survey data in development economics include social desirability bias (e.g., Bertrand and Mullainathan, 2001; Zwane et al., 2011), sampling bias (e.g., Mathiowetz and Groves, 1985; Bardasi et al., 2011; Serneels et al., 2016) and the Hawthorne effect (e.g., Zwane et al., 2011; Smits and Günther, 2018).[1] Furthermore, several social science studies find that respondents whose true values are large tend to underreport, while those whose true values are small tend to overreport, leading to mean-reverting measurement error, which has been observed in the reporting of labor market outcomes (Bound and Krueger, 1991; Bound et al., 1994; Pischke, 1995; Bound et al., 2001; Bonggeun and Solon, 2005), Body Mass Index (O'Neill and Sweetman, 2012), educational attainment (Kane et al., 1999), and in a study on consumption in developing countries (Gibson et al., 2015). All of these problems can create systematic errors and thus, reduce measurement accuracy and bias regression coefficients in econometric analysis. In experiments, these errors are particularly problematic if they have varying effects

* Corresponding author.
E-mail address: adina.rom@nadel.ethz.ch (A. Rom).

[1] The extent to which the Hawthorne effect influences social science research has been widely debated (Adair et al., 1989; Leonard and Masatu, 2006; Levitt and List, 2011; Clasen et al., 2012; McCambridge et al., 2014).

https://doi.org/10.1016/j.deveng.2020.100056
Received 10 October 2019; Received in revised form 6 August 2020; Accepted 10 August 2020
Available online 28 September 2020
2352-7285/

across treatment groups. In addition, respondents may simply recall answers incorrectly (Bertrand and Mullainathan, 2001; Beaman et al., 2014). Such recall errors seem to increase as time between the event or behavior and the survey passes. But even if the time between an event and a survey is short, data might still be noisy for anything that fluctuates substantially over time (e.g., incidents of diarrhea), even if the population mean is accurately estimated. In many cases, collecting high frequency data for such events is nearly impossible because it is often intrusive, expensive, and logistically challenging. These sources of random error do not necessarily lead to systematic error, however, if the dependent variable is affected, these errors can reduce the precision of estimates. If the explanatory variables are affected, this can lead to an attenuation of coefficients towards zero. Hence, random errors reduce the chances of detecting an effect of a new policy or technology or identifying differences between sub-groups. Moreover, these errors can still lead to systematic biases if they are more pronounced for certain sub-groups. Loken and Gelman (2017) even argue that random measurement errors can increase the chances of finding spurious correlations in small samples.

Recognizing these challenges, development economists have begun comparing different types of survey questions and methods. Typically, these studies aim to measure the extent of the problem and to optimize survey tools. A number of studies analyze recall biases in surveys. For example, Das et al. (2012) and Beegle et al. (2012) study the optimal length of recall times, while others compare recall answers with diaries (Deaton and Grosh, 2000; De Mel et al., 2009) or analyze whether asking aggregated questions versus disaggregated questions leads to more accurate and precise estimates (Grosh and Glewwe, 2000; Daniels, 2001; De Mel et al., 2009; Arthi et al., 2016; Serneels et al., 2016; Seymour et al., 2017). These studies, however, often compare different types of self-reported data to each other or they compare self-reported survey data with administrative records. Thus, they tend to rely on benchmarks whose accuracy is also unclear.

Prices for sensor technology have dropped significantly in recent years and more “off-the-shelf” solutions have become available (IPA, 2016; Pillarisetti et al., 2017), meaning sensors can now be used to collect data in studies with larger sample sizes. Although sensors present a different set of potential measurement limitations (e.g., technological failure, see Section 2.3 for more details), they provide new opportunities for researchers to avoid some of the problems posed by survey data and represent a new benchmark for survey data. Sensors are particularly suitable for studying the adoption of new technologies to improve the lives of poor households, such as water filters, cookstoves, or, in our case, solar lights. The use of such technologies cannot be measured with remote sensing data as they are frequently used within the house. Other feasible measurement technologies, such as video footage, are very invasive. In contrast, sensors can be easily attached to household devices without interfering with use.

In this study, we use data from 220 sensors[2] and a corresponding household survey to describe patterns of small, solar light usage, that households received for free or had the opportunity to purchase (August 2015–March 2016).[3] Sensors logged whenever the solar light was switched on or off, providing high frequency usage data for households across the study period. As Rom and Günther (2019) show, switching to renewable energy sources and more energy efficient appliances can have important health and environmental benefits, however, households only realize these benefits if they actually use the solar light and reduce the use of kerosene accordingly. Even very promising technology can fail to be effective because it is simply not used (e.g., Hanna et al., 2016). Therefore, it is crucial to get an accurate understanding of households' solar light use patterns to estimate the effect of this technology.

We compare sensor data with survey data, interviewing two different household members from each household. The interviews included both detailed (time diary) and global household questions about solar light use, allowing us to learn about social desirability bias, selection bias, mean-reverting measurement error, and random error in survey data. In addition, our experimental set-up allows us to get an indication of the magnitude of the Hawthorne effect, since the survey team visited a random sub-sample of respondents more frequently during the first two months of the study.

The main findings of our study are first, that — according to sensor data — households use the solar light for 3.9 hours per day, on average. About 60% of households use the solar light every single day. In contrast to much of the literature using sensors to study technology adoption in developing countries, we do not find systematic overreporting of usage: the averages of survey data and sensor data look fairly similar. However, consistent with mean-reverting measurement error, we find that households that hardly used the solar light tend to overreport use. Households that use the solar lights frequently, on the other hand, tend to underreport use. Third, we find that more frequent household visits from surveyors increased use of solar lights initially, but had no effect in the long run when visits stopped. Hence, the Hawthorne effect only biased use in the short term. Fourth, at the household level, there is little correlation between the daily light use estimated with diary questions and the sensor data, while the correlation is higher when using total estimates of household usage for the previous day. Finally, increased precision of the sensor data allows us to see usage patterns of sub-groups more clearly in comparison to survey data. Sensor data reveal that poorer households tend to use solar lights more often.

Our paper is related to the small, but burgeoning body of research that uses sensor data to understand technology adoption in low- and middle-income countries. Some of these studies also compare sensor data to survey data, in particular, studies measuring cookstove use (Ruiz-Mercado et al., 2013; Thomas et al., 2013; Wilson et al., 2016; Ramanathan et al., 2016; Piedrahita et al., 2016).[4] In contrast to our findings, all but one of the studies (Piedrahita et al., 2016) find little correlation between self-reported use of cookstoves and sensor data and suggest that survey data significantly overestimate cookstove adoption. It is likely that respondents overreport cookstove use because they think the adoption of the new technology is socially “desired” given that using it has positive externalities.[5] Our study may differ from this literature

[2] IPA (2016) defines a sensor as a “device used to measure a characteristic of its environment— and then return an easily interpretable output, such as a sound or an optical signal. Sensors can be relatively simple (e.g., compasses, thermometers) or more complex (e.g., seismometers, biosensors).”

[3] The full study was ten months long, however, here we present eight months worth of sensor data, since August is the first month in which every household had the solar light for a full month.

[4] In a field experiment in Guatemala, Ruiz-Mercado et al. (2013) used stove use monitors to study the use of improved cookstoves for 32 months. Wilson et al. (2016) studied cookstoves in Darfur for 1–3 months. Ramanathan et al. (2016) studied usage in rural India for 17 months. Piedrahita et al. (2016) studied cookstove stacking by monitoring multiple cookstoves with sensors and survey data in Northern Ghana for 12 months. In a field study in Rwanda, Thomas et al. (2013) compared reported usage of cookstoves from monthly surveys with sensor data from the same respondents over the course of five months. These studies on improved cookstoves seem highly relevant for comparison for a number of reasons. First, cooking and lighting typically represent the most urgent energy needs of rural households in low-income settings. Second, similar to solar lights, improved cookstoves currently receive a lot attention from international donors and policy makers: the hope is that these technologies can improve human health and reduce environmentally damaging emissions. Finally, the effectiveness of both technologies depends to a large extent on whether households replace the use of the “old” technology (i.e., the old cookstove or kerosene lanterns) with the “new” technology (i.e., an improved cookstove or solar lamps).

[5] Examples of other socially-desired technologies include water filters (Thomas et al., 2013), latrines (Garn et al., 2017; Gautam, 2017), vaccines (Banerjee et al., 2010), and bed nets (Cohen and Dupas, 2010).

because we observed much higher rates of solar light adoption than the cookstove adoption reported by these studies and because cookstoves tend to be used by particular household members at specific times of day, whereas solar lights could feasibly be used by all household members at any time of day.

In this context, our findings have a number of implications for survey and sensor measurements. First, the added value of sensors seems to be particularly high when technological devices are used by several people within the household and when biases in survey data are expected to be large. For the case of household technology adoption, our results along with previous studies imply that social desirability bias seems to be a challenge for technologies that require behavioral change (cookstoves), while mean-reverting error is a challenge for technologies that are adopted quickly. Sensors also complement survey data well when high frequency data adds a lot of value or when precise estimates are needed to answer the research question, such as if the sample size is small or sub-group analysis is important.

Second, for surveys, our data suggest that asking about global use estimates provides more accurate results than asking two household members about their individual use throughout the day (time diary) and combining them. Thus, while time diary questions are relevant for understanding use patterns over the course of the entire day, they do not seem to be ideal for understanding global use of a device that is shared by many household members. Third, we find that frequent interactions with field staff can temporarily increase use of new technologies, suggesting that researchers need to think carefully about how interactions with the field staff could bias results and, if this is a concern, find ways to measure these surveyor effects.

Finally, sensor attrition raises important questions about whether attrition is correlated with usage, however, we do not find clear evidence that this is the case. Due to sensor attrition, we focus most of our analysis on measurements taken in the first month.[6] Beyond taking steps to minimize attrition (e.g., thorough pilot testing), studies using sensors may require adaptive protocols that account for sensor attrition or malfunction of the studied technology.

The remainder of this paper is structured as follows. The next section describes the research setting, a solar light intervention in rural Kenya, and the sensor and survey data used to measure solar light adoption. Section 3 describes solar light usage patterns using the sensor data. Section 4 compares sensor data with survey data and studies to what extent survey data might be limited by social desirability bias, mean-reverting error, sampling bias, and random measurement error linked to recall errors when analyzing household technology adoption. Section 5 concludes.

2. Study design and data

The sensors used in this paper are part of a larger randomized controlled trial (RCT) conducted between June 2015 and March 2016 in two sub-counties in Busia, western Kenya. The sample contained 1,410 randomly selected households from a random sample of 20 schools (i.e., 70 households per school). To enter the pool of potential households, a household had to have a student in class five, six, or seven in one of the 20 randomly selected schools. Randomization into different treatments was conducted at the household level and stratified at the school level. In total, 400 households were randomly assigned to the control group, 400 received a solar light for free, and 209 households received an offer to buy a solar light at 900 KES (US $9), 201 households at 700 KES (US $7), and 200 households at 400 KES (US $4). In each household, we surveyed the respective student and that student's caretaker.

Half of the households that received a free solar light were given a Sun King Eco and half received a Sun King Mobile light (see Appendix A, Figs. A.1 and A.2 for pictures), both manufactured by Greenlight Planet and quality assured by Lighting Global, a joint initiative of the World Bank and the International Finance Cooperation. At the time of the study, the Sun King Eco sold for US $9 in Kenya and the Sun King Mobile for US $24. The lights require between five and eight hours to fully charge. According to tests conducted by Lighting Global, the Sun King Eco provides light for 5.8 hours when used at its maximum brightness of 32 lumens. The Sun King Mobile can be used for 5.4 hours on its brightest mode (98 lumens) and can also charge a mobile phone (Lighting Global, 2015; Greenlight Planet, 2016). The lights last longer if used at lower lumen levels (the lights have three brightness modes) and for a shorter period of time if they are also used to charge a mobile phone. For comparison, a simple tin lamp, which is what was used most often for indoor light in our sample, provides around 7.8 lumens and a kerosene lantern provides 45 lumens (Mills, 2003). Thus, both types of solar lights provide much stronger light than the tin lamp if used at their maximum brightness.[7] The 610 households that randomly received a voucher to purchase a solar light were all offered the opportunity to buy the Sun King Eco model.

Households were encouraged not to give away or sell their solar lamp to other households. To understand if this was a problem, when surveyors visited households 3–4 months into the study for the second sensor data collection (see the project timeline in Fig. 1), they asked to see the solar light. Only 3.6% of respondents were not able to show their solar lights. Of the 400 solar lights that were distributed for free to households in June and July 2015, 164 were equipped with a sensor that measured usage. Households only learned about the sensors when we asked for permission to download their data for the first time in July 2015 (see Fig. 1), which was a few weeks after the baseline survey. The research team only accessed the data if the respondent gave permission for them to do so. All households gave us permission to download the data. Of the 130 solar lights that were sold to households in June and July 2015 at either 900 KES (US $9) or 700 KES (US $7), a sub-sample of 56 solar lights was equipped with a sensor that collected data.[8] Thus, in total, we had 220 solar lights equipped with a functioning sensor (see Section 2.1). A random sub-sample of those 220 lights with a sensor (37.1%) were subject to around five additional household visits in August 2015. Other studies have found that more frequent interactions between households and surveyors led to increased use (Zwane et al., 2011; Wilson et al., 2016). We use the variation in visits in our study to also analyze whether additional household visits led to more solar light use.

Activity	Date
Baseline Survey	June/July 2015
Free Solar Lights/Voucher Distributed	June/July 2015
First Solar Sensor Data Collection	July 2015
Sensor Follow Up (random sub-sample visited)	August 2015
Second Solar Sensor Data Collection	September/October 2015
Endline Survey	February/March 2016
Third Solar Sensor Data Collection	February/March 2016

Fig. 1. Project timeline.

[6] 6.8% of sensors stopped functioning within the first month of the study and 37.7% before the end of the eight-month study period (for comparison, survey response attrition at the end of the study was 5.9% for adults and 9.1% for pupils). It was difficult to confirm why sensors stopped functioning without potentially damaging a respondent's light, however, we know the most likely reasons for attrition are that the sensor simply malfunctioned or the light broke and disabled the sensor.

[7] In the analysis we combined both types of solar lights as we did not observe significant differences in usage patterns.

[8] A total of 610 households received an offer to buy a lamp, but only 274 bought one. Out of these, 130 were sold at either 900 KES (US $9) or 700 KES (US $7) and the remaining 144 were sold at 400 KES (US $4).

In the beginning of our study, only 4.2% of the sampled households had access to some form of electricity, with only 1.4% of households connected to the grid, 1.1% with access to a solar home system, 1.5% with access to a car battery, and 0.1% with access to a generator. Most of these households were using the respective electricity source for their radio (80.0%), for lighting (72.0%) or to charge their mobile phones (65.4%). Just under a third of people with access to electricity used it to watch TV. No one had a refrigerator and no one used the energy source for activities that are potentially income generating, such as sewing, water pumping, or irrigation. Most households (88.4%) relied on small locally produced kerosene lights (tin lanterns) for lighting, while others used larger kerosene lanterns (5.3%) or solar lights (3.8%) as their primary lighting source. On average, a household owned 2.1 tin lamps and only 6.5% owned a solar light at baseline. Every household that used grid electricity also used at least one other source of lighting — probably a reaction to the frequent blackouts in the study region. Moreover, households in our sample were generally large (more than six household members) and very poor, with 84% living on earth floors and 62% frequently having to cut meals (see Appendix B, Table B.1).

2.1. Sensor data

We used Bluetooth-enabled Solar Lamp Usage Monitors (referred to as sensors or solar sensors throughout this paper) to determine when the lamp was in use by measuring the change in voltage of the solar lamp's light emitting diode (LED).[9] The sensor was installed by soldering the sensor to the circuit board inside the light. Using smartphones enabled with Bluetooth and an iPhone application ("Lamplogger"), field officers visited households and wirelessly uploaded data directly from the sensor to the phone. Respondents first became aware their light had a sensor installed inside it about one month after they received the light, when our field team visited to collect data for the first time (see Fig. 1). The sensors, along with the iPhone application, were specifically developed for this study. The data on the sensors could only be accessed via this specific iPhone application. It was, therefore, impossible for households to download or check their own usage data. Since the use of sensors in field experiments is still relatively new and other researchers may find themselves in a similar situation to ours, we share a few key lessons learned about implementing and managing sensor technology in the field in Section 2.3 and in Appendix C.

Field team members visited households to collect sensor data in July 2015, between September and October 2015, and between February and March 2016; hence, about one month, three months, and seven months after light distribution, respectively (see Fig. 1).[10] We have sensor data for a total of 220 households for at least part of the eight-month study period. However, by the endline survey (February–March 2016), around a third of sensors had stopped recording data, such that we were left with 147 sensors (see also Appendix A, Fig. A.3).[11] Sensor attrition can have several reasons. First, we were not able to find five households with sensors during endline data collection. For the remaining 68 sensors, the sensors stopped recording data either because the battery died, the sensor was faulty (manufacturing error), or the solar light stopped working. While, unfortunately, we cannot deduce from the sensor data which of these issues occurred, 29 of the 68 households with a sensor that stopped working before the end of the study indicated during the endline survey that their solar light stopped working, while 39 indicated that the solar light still worked. This sensor attrition rate of 30.1% is similar to the failure rate Thomas et al. (2013) observed with sensors applied to monitor water filter use, but higher than the sensors they used for cookstoves (18%) and also higher than what Wilson et al. (2016) and Ruiz-Mercado et al. (2013) found in their study with 17% and 10%, respectively. It is possible that the point in time at which the sensor stopped working is correlated with usage. On the one hand, it could be that some sensors may have stopped working because the solar light was not used for a number of consecutive days. On the other hand, it is also likely that solar lights that are used more intensively tend to break more often. However, if, for every month of the study, we compare the usage of lights that broke in the previous month with lights that did not break, the coefficients go in different directions (see Appendix B, Table B.2). Therefore, it does not seem that one effect dominates the other. We also looked at correlates of sensor attrition before the end of the first month (August) and before the end of the study (Appendix B, Table B.3 and B.4). While most household characteristics were uncorrelated with sensor attrition, sensors in larger households and sensors in households without access to modern energy were slightly more likely to stop working. For these reasons, it is possible that we under- or overestimate usage when using data from the end of the study. To avoid possible biases in sensor measurements, we focus most of our analysis in Section 3 and 4 on the first month of sensor data collection only, when 93.2% of the sensors were still working by the end of the month. We replaced data points with missing values once the sensor stopped logging data. In this sense, all results should be interpreted as "usage conditional on the lights functioning."

For the sensor data, we report the following measures of average daily solar light use:

- Entire Study (all): recorded average use by day and sensors, no matter how long they worked (N = 220). Data were used from all the days for which we have data. Once a sensor stopped working the remaining days were coded as missing. Months included: August 2015–March 2016. Variable: Sens (All)
- Entire Study (worked entire study): sensors that worked until the end of the study (N = 147). Data were used from all the days for which we have data. Months included: August 2015–March 2016. Variable: Sens (All) worked until End
- First Month (all): recorded average use by day and sensors, no matter how long they worked (N = 220). Data were used from the first month of the study. Once a sensor stopped working the remaining days were coded as missing. Month included: August 2015. Variable: Sens (Aug)
- Previous Day: sensors that worked until the end of the study (N = 147). Data were used from the day before endline data collection. Days include: Varying days in February and March 2016, depending on the day the endline was conducted in each household. This measurement is used to compare sensor to survey data in Section 4, since we asked about solar light use on the previous day in the survey. Variable: Sens (Yest.)

Sensors tracked when the solar lights were turned on and off. Based on this information, we calculated the total number of minutes a solar light was used on any given day of the study. Independent of the measure used, we first calculated average daily use by sensor, meaning that we always weight each sensor equally, regardless of the number of days of data we have.

2.2. Survey data

The survey data refers to the endline survey, which was conducted between February and March 2016 (see Fig. 1) and contained, among

[9] Sensors were developed by Bonsai Systems: https://www.bonsai-systems.com.

[10] We collected data multiple times throughout the course of the study in order to check on sensor attrition and other technical problems. In theory, however, we could have collected data only once at the end of the study and retrieved the exact same data. Please note that the fact that we present here four types of measurements and that we have four instances of data collection is purely a coincidence.

[11] We ended the study slightly before the end of March when there were still 147 surviving sensors. The figure in Appendix A shows the total number of working sensors at the end of March, hence, why the totals are slightly different.

others, questions about household light use habits (the full survey is available from the authors upon request).

Information about solar light use came from two separate questions:

- Detailed Questions (see Appendix D): a separate battery of questions asked each individual about their activities and light use[12] for half-hour long time slots between 3 a.m. and 7 a.m. and between 6 p.m. and 11 p.m. of the previous day (9 hours total), corresponding to nighttime (dark) hours in Kenya. We faced a trade-off between level of detail (including daytime hours) and survey length. Ultimately, we only asked for the most detailed information about light usage at night in order to limit both financial costs and the opportunity cost to respondents in terms of patience and attentiveness (N = 215 for adults, N = 205 for children). From these questions we created two variables:
  - We created a dummy indicating whether the adult used the solar light during each time slot in the time use section and then aggregated all relevant time slots to get the total number of hours used. Variable: Surv (Detail) Adult[13]
  - We created a dummy indicating whether either the adult or the child or both used the solar light during each time slot in the time use section and then aggregated all relevant time slots to get the total number of hours used by one or both of the respondents. Variable: Surv (Detail)
- Aggregated Question (see Appendix D): one question asked the adult respondent for an estimate of total solar lighting used by the household on the previous day (N = 161). This question was only asked if the respondent indicated that they had a functioning solar light. Variable: Surv (Aggr.)

It is important to note that a respondent was only asked the Aggregated Question if they indicated that "any of their solar lights still works," due to a skip pattern in our survey instrument. A total of 53 households reported that their solar light did not work, however, of these 53 households, 21 (40%) still had a working solar light and had, according to the sensor data, used it the previous day, suggesting that either they did not understand the question, did not know that their light still worked, or intentionally deceived the surveyors.[14] Thus, we only have an answer to the Aggregated Question from 161 of the households with sensor data.

The Detailed Questions were not affected by this skip pattern, as we asked these questions of both adults and pupils as part of the time use section of the survey, thus all households were asked. In this section, it was not obvious to the respondent that the questions were about use of solar lights as the focus was on their activities for each half-hour of the day. The Aggregated Question was asked towards the end of the survey as part of a module on solar light use. We placed the questions about solar lights later in the survey to avoid priming respondents for the other sections of the survey.

2.3. Advantages and disadvantages of sensor data

Sensors and the data they collect can have several advantages over survey data, allowing researchers to collect high frequency information about the use of a technology over an extended time period. Such information is usually very time consuming and intrusive to collect with surveys, especially if a technological device is used by several people, who all need to be individually and repeatedly asked about the timing of their usage. For example, in our case, the adults we interviewed simply might not know whether their children used the solar light at night. One would have to separately ask all household members to get the full picture. In addition, asking respondents about events or behaviors that lie in the past might lead to very noisy and perhaps even biased results.

Sensors, though not susceptible to random measurement error, sampling error, Hawthorne effects, mean-reverting error or social desirability bias, have their own limitations. One limitation of the sensors used in this study is that they cannot distinguish between users, so inequality in technology access within the household cannot be estimated. In addition, when sensors stop recording events, it is not always possible to understand what went wrong. For example, it is not easy to determine if the source of the problem is the light not being used, the light malfunctioning, or the sensor malfunctioning. In addition, once a sensor breaks, nothing more can be said about use of the solar lights over the period of time for which data is missing, whereas survey data can still be collected even if the sensor breaks. There is also a higher risk of data loss when using sensors that cannot be tracked remotely. In the case of our sensors, if the sensor broke between two data collection rounds any data not already collected was lost.

Finally, researchers might underestimate the trade-offs between sample size and study duration on the one hand and data collection and management costs on the other hand. First, while sensors are a more cost-efficient means of studying frequent behavior over longer study durations, current sensor technology, at least, is not yet useable in studies with very large sample sizes. Second, while the data collection itself is much cheaper when compared with survey data collection, managing sensors and solving problems that affect many households over a long period of time is costly. Managing sensors and troubleshooting problems requires considerable management and field staff time and sometimes necessitates more visits to the sensor than planned, increasing concerns about the Hawthorne effect. Field staff also need considerable extra training on handling sensors and a technician is often needed. More lessons learned for researchers on how to manage sensors for data collection can be found in Appendix C.

3. Use of solar lights

Sensors can provide detailed information on how usage of a technology varies throughout the day, the week, the month, and the year. As discussed in Section 2.1, we focus on results from the first full month of the study (August 2015) for the analysis of solar light use, since about 93% of the sensors worked through August, whereas by March 2016, an additional 13.6 sensors had dropped out each month on average (see Appendix A, Fig. A.3). That said, results for the entire study period are very similar to results from the month of August (results for the entire study period are available from the authors upon request).

3.1. Solar lights are used frequently, mostly between 7:30 p.m. - 8:30 p.m

Households used the solar light on average 6.4 out of seven weekdays and 58.6% of households used the solar light on every single day of the study. Households used the solar light for 3.86 hours per day and 71% of households used the solar lights between two and five hours per day (see Fig. 2 and Table 2, Row 3). Daily use across the entire study period is actually slightly higher (4.07 hours per day), possibly since schools were still closed during the first two months of the study (Table 2, Row 1). There are only nine households (4% of all households with sensors) who

[12] Options: Electricity-powered lighting, Solar home system powered lighting, Tin Lamp, Kerosene lantern/Hurricane, Fire, Wood, Battery-powered torch/lantern, Candle, Solar lantern/solar torch, Pressurized Kerosene Lantern, Other rechargeable lantern, Cell phone light, No lighting used, Matchsticks, Other.

[13] We used the following equation to calculate use:

$$y(x) = \frac{\sum_{t=1}^{T} \mathbb{1}\{x_t = \text{used solarlight}\}}{2} \quad (1)$$

where x are half-hour slots between 3 a.m. and 7 a.m. in the morning and 6 p.m. and 11 p.m. in the evening of the previous day with t = 1,2, …, T and T = 18. y can hence take values between 0 and 9.

[14] According to sensor data, households which indicated that at least one of their solar lights worked during endline did not use their solar lights for different amounts of time per day than households that said that none of their solar lights worked.

used the solar light for less than one hour per day on average (Fig. 2). These findings of high rates of solar light usage across all households contrast with recent findings about improved cookstoves. Wilson et al. (2016), for example, find that 29% of households hardly used the technology.[15]

As explained in Section 2.3, sensors allow researchers to collect high frequency data. Fig. 3 shows the share of solar lights that were used, reported in half-hour slots, averaged over all days of the first month of the study. We created a dummy for every half-hour slot, which equals one if the solar light was used for more than 15 min in a row during that half hour and zero otherwise. We then calculated, for each sensor, the percentage of days that the light was on (as a percentage of all days that the sensor worked in August) and used this information to calculate the average across all sensors. We find that households mostly use the solar light during evening hours. The half-hour intervals when most solar lights (81.94%) were switched on was between 7:30 p.m. and 8:30 p.m., which is right after sunset in Kenya. As expected, there is also a peak, albeit a smaller one, during morning hours, in particular between 6:00 a.m. and 6:30 a.m. Interestingly, between 15% and 20% of households also have the solar lights switched on during nighttime hours. Anecdotal evidence suggests that, among other reasons, some use the solar light as a security light during the whole night or when they get up to use the restroom or check on their cattle. As expected, use is lowest during the day — only 1.05% used them during the daytime (between 9:00 a.m. and 5:00 p.m., see Fig. 3).

On average, households switched the light on and off 4.74 times per day (SD 3.35) with each on/off event lasting an average of 50.71 minutes (SD 93.32); 50% of all use events were shorter than 12 minutes. In theory, households could leave the solar lights on all the time, also during charging, which would make sensor measurements meaningless. However, there are only 11 households that used it for more than 8 hours per day over the study period. Checking when these households used the light, we see that these high-usage households used the lights more during the night, and not during the day.

3.2. Usage does not vary across months, but is lower on weekends

Sensors can also be used to study changes in use over time. Households might increase use of a product as they learn about its advantages or develop a habit of using it. Households might decrease use if they discover unexpected disadvantages or if their excitement over the novelty of the product wears off over time. Use could also vary with the schooling or agricultural schedule. Fig. 4 shows use over the eight months of the study period for the 147 solar lights for which we have data until the end of the study. Use was slightly lower in August and September, but none of the differences are statistically significant (Appendix B, Table B.5). This pattern could be linked to the fact that schools were closed in August, due to holidays, and in September, due to a teacher strike. However, as explained in Section 2.1, around one third of the sensors did not survive until the end of the eight-month study and we do not know how use would have evolved amongst those households whose lights/sensors did not survive. There is no clear pattern indicating whether sensors in high-usage solar lights were more likely to stop recording data than sensors of low-usage solar lights (see Appendix B, Table B.2). Fig. 4 also breaks down usage by day of the week. We observe that solar lights are used less on the weekend. This difference is statistically significant at the 5% level (Appendix B, Table B.6).

3.3. Intense monitoring increased use temporarily

A random 37% of the sampled households with solar lights and sensors were exposed to more frequent visits by surveyors at the beginning of the study (during August 2015, see Fig. 1). More frequent visits did increase use in August 2015, however, this difference disappeared quickly thereafter — already in the second month of the study, when visits stopped (Table 1). Different mechanisms might explain this difference: respondents might have felt more observed and used the novel product more as a result (Leonard and Masatu, 2006; Clasen et al., 2012), the visits may have made the product more salient, i.e., reminded respondents of the product (Zwane et al., 2011; Smits and Günther, 2018), or the surveyors might have accelerated learning about the product.

4. Comparing survey and sensor data

In this section, we analyze whether estimates of technology use based on survey data differ from those obtained from sensor data (Section 4.1). Moreover, we test several hypotheses that have been intensively discussed in the literature that deal with bias in survey data (Section 4.1-4.3). Lastly, we analyze whether sensor data, which measure technology use with higher precision, allow us to detect differences across sub-groups or experimental treatments with smaller sample sizes (Section 4.4).

4.1. Averages from sensor and survey data are similar

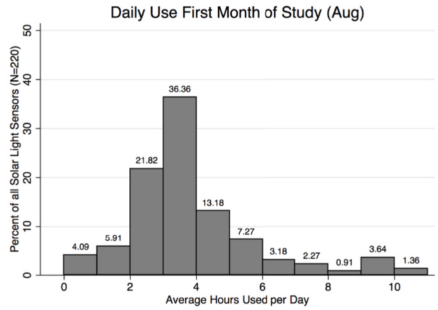
Comparing the three different survey measures (see Section 2.2) with the four sensor measures (see Section 2.1), we find that the averages from the sensor data and from the survey data are relatively similar (Table 2). If anything, survey data suggest a slightly lower use of solar lights than sensor data (see also Table 6 and Section 4.3). This finding stands in contrast to most of the recent literature (Thomas et al., 2013 or Wilson et al., 2016, for example) studying the use of improved cookstoves with sensor and survey data, which finds that respondents tend to overreport use on average. There are, however, two important differences between our study and previous work. Namely that, in our case, adoption of solar lights was high (see Section 3.1), while adoption of improved cookstoves was typically low. Moreover, the solar light is a technological device being used by many household members, whereas a cookstove is typically only used by a few.

A second interesting finding is that all sensor measures — whether looking at the first month, the entire study period, or yesterday — reveal very similar solar light usage (differences in means are not statistically different from each other at the 5 percent level). Hence, solar light usage does not seem to fluctuate much over time (see also Appendix B, Table B.5) and attrition of sensors (see Section 2.1) does not seem to be correlated with high or low usage. This result also indicates that asking survey questions about the previous day would, in theory, be a good estimate of a specific household's average solar light use: unit-level survey estimates from the previous day should not diverge largely from sensor measurement averages over a longer time period (as would be the case, for example, for diarrhea estimates).

4.2. Frequent users underreport, infrequent users overreport

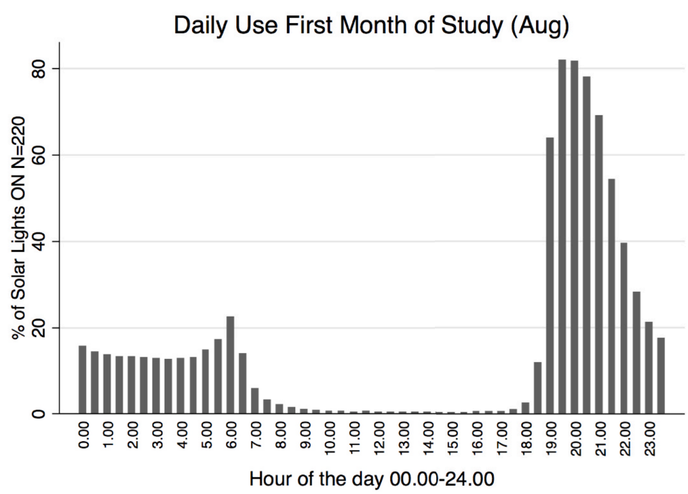
Even if averages of sensor and survey data are similar, systematic measurement error can still exist if measurement error is correlated with household characteristics that cancel out at the mean or if mean-reverting measurement error is present. Mean-reverting measurement error means that measurement error is negatively correlated with the true value (Bound and Krueger, 1991). Fig. A.4 in Appendix A indicates that households that hardly use the solar light tend to overreport use (difference between sensor measurement and survey measurement is negative), while households that use the solar light a lot tend to underreport use (difference between sensor measurement and survey measurement is positive). This so-called mean-reverting measurement error has also been shown by Gibson et al. (2015) for consumption data collected using household surveys in low-income settings. However, the benchmark for various survey measures in the study was also a survey

[15] They defined "non-users" as those using the cookstove less than once on 10% of days.



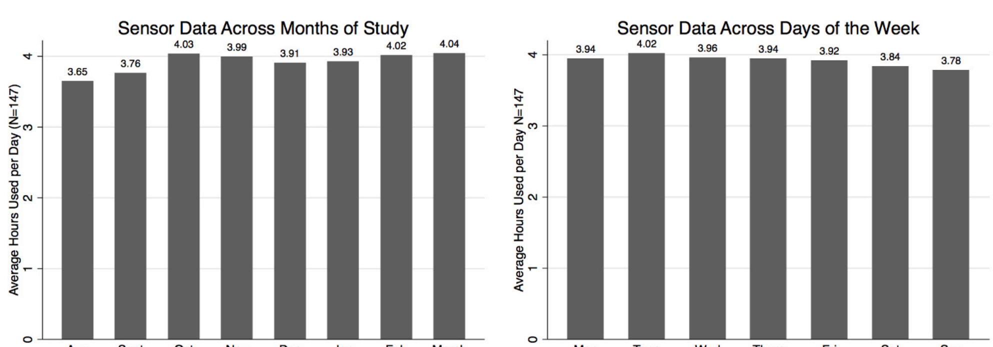
Notes: This graph shows sensor data on the average number of hours the solar lights were used per day during the first month of the study.

Fig. 2. Average hours solar lights are used per day.



Notes: We classify usage by whether the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot. We then calculated for each sensor the percentage of days that the light was on across all days that the sensor worked and then used this information to calculate the average across all sensors.

Fig. 3. Use across the day.



Notes: Sample is restricted to sensors that worked until the end of the study.

Fig. 4. Daily use across months of the study and across days of the week.

measure (individually-kept diaries with daily supervision over 14 weeks). Therefore, the benchmark in Gibson et al. (2015) might be vulnerable to systematic and random measurement error itself (see Section 3.3 and 4.3), whereas in this paper, we benchmark survey data against sensor data — which, of course, has its own shortcomings, especially over time, but less so for a single day or short periods of time (see Section 2.3).

To formally test for mean-reverting measurement error, we follow the methodological approach proposed by Bound and Krueger (1991). We first compare the variance of survey measures with the variance of

Table 1
Hawthorne effect.

VARIABLES	(1) Sensor (Hrs) First Month	(2) Sensor (Hrs) First Month	(3) Sensor (Hrs) All Months	(4) Sensor (Hrs) All Months
Frequent Visits	0.584**	0.589**	0.339	0.278
	(0.284)	(0.296)	(0.253)	(0.239)
Observations	220	147	220	147
R-squared	0.019	0.026	0.009	0.009
Mean	3.646	3.285	3.941	3.629

Notes: Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Column 2, 4 and 6 are restricted to those sensors that worked until the end of the study. The remaining columns show every household for which we have at least one data point during the relevant time period. Column 3 has fewer observations since in the beginning of September only 205 sensors remained (see Appendix A, Fig. A.3). No control variables were used.

Table 2
Mean light use (Hrs) per day: Survey and sensor data.

	(1) All Data Mean (SD)	(2) All Data Obs	(3) Exclude Missing Means (SD)	(4) Exclude Missing Obs
(1) Sens (All)	4.067	220	3.813	125
	(1.776)		(1.464)	
(2) Sens (All)- Worked until End	3.731	147	3.813	125
	(1.404)		(1.464)	
(3) Sens (Aug)	3.864	220	3.607	125
	(2.031)		(1.846)	
(4) Sens (Yest.)	3.706	147	3.777	125
	(2.132)		(2.247)	
(5) Surv (Detail)	3.388	215	3.616	125
	(1.764)		(1.625)	
(6) Surv (Detail)- Adult	3.193	215	3.152	125
	(1.377)		(1.371)	
(7) Surv (Aggr.)	3.573	161	3.492	125
	(2.073)		(2.030)	

Notes: Column 1 and 2 include all data, Column 3 and 4 only the 125 observations where we have all sensor and survey variables listed in this table (see Section 2 for further explanations). Row 1 includes all sensors no matter when they stopped working, Row 2 includes data from all sensors for the month of August only, Row 3 includes sensor data for the day before the study, Row 4 shows survey data from the Detailed Questions for adults and pupils combined, Row 5 shows the same question as Row 4, but only for adults, and Row 6 shows the Aggregated Question where we asked about use of the entire household (see questions in Appendix D). Note that the survey questions refer to the day before.

sensor data. If measurement errors in survey data are random then the variance of the survey measures should always be higher than the variance of sensor data, however, this is not the case in our data (see Tables 3 and 4, column 4). In a second step, we regress the sensor measurements (the benchmark) on the survey measurements. If measurement error is not mean-reverting the coefficient should be one. If mean-reverting measurement error exists, then this coefficient will be less than one. We obtain a coefficient that is significantly smaller than one for all survey measures (see Tables 3 and 4, column 5). The survey measurement based on the global (Aggregated) question about solar light use is thus, not only closest to the sensor measurement on average (Table 3, column 1), it is also the survey question with least mean-reverting measurement error (Table 3, column 5) and shows the closest correlation with sensor data at the unit level (see Section 4.3).

There could be a couple of explanations for this observation. First, respondents could have a certain reference point in mind regarding reasonable light use that they report regardless of actual light use. It is also possible that underreporting occurs because respondents are not aware of other household members' use (especially in high-usage households), while respondents who hardly use the solar light over-report because they feel they are expected to use the light (social desirability bias).

We obtain similar results when we use the sensor measurement of "yesterday" as the benchmark (Table 3) or the sensor measurement of the "first month" of the study (August 2015) as the benchmark (Table 4). Note that in contrast to Table 2, where we showed all possible sensor measurements that can be derived from our data set, we now (and in the following section) focus on the sensor measurement of "yesterday" and the "first month" of the study. We focus on these measures for three reasons. First, sensor measures do not deviate much from each other (Table 2), second, "yesterday" is directly comparable to survey data (given that both measure solar light use on the day before the survey took place) and third, the "first month" of the study has the most data points with minimum sensor attrition.

Analyzing other correlates of measurement error, such as design variables (free vs. purchased solar light and more frequent visits) and various household characteristics (type of floor, food security, wealth index, education level, household size, and energy access), we see that households with access to modern energy are more likely to underreport use than households without access. However, this difference is driven by only 10 households who had access to modern energy sources and a sensor that worked until endline (Appendix B, Table B.7).

4.3. Detailed Questions less correlated with sensor data than aggregated question

In the survey, we asked about solar light use in two different ways. First, we asked adults and children to report the activities they engaged in for each half-hour slot between 6:00 p.m. and 11:00 p.m. and between 3:00 a.m. and 7:00 a.m. and whether they used any lighting for each activity and time slot. Second, we asked adults to estimate the global use of solar lights by the entire household on the previous day (see Section 2.2 for more details).

Using sensor data, we calculated the percentage of days that the light was used during that specific time slot for each sensor (across all days that the sensor worked), and then used this information to calculate the average across all sensors. By "used" we mean that the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot.

In Figs. 5 and 6, we compare the estimates based on the Detailed Questions with the sensor measures. Overall, we see that the patterns of solar light usage over the course of the day match well. Note that in the

Table 3
Tests for mean-reverting measurement error - yesterday.

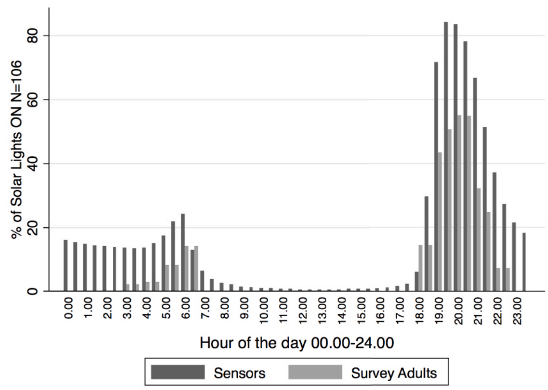
	(1) Mean	(2) Ratio to benchmark (Means)	(3) Variance	(4) Ratio to benchmark (Variance)	(5) Beta (SE)	(6) P-Val
(1) Surv (Detail)	3.388	0.914	3.111	1.459	0.048 (0.089)	0.591
(2) Surv (Detail)- Adult	3.193	0.862	1.896	0.889	0.080 (0.074)	0.285
(3) Surv (Aggr.)	3.573	0.964	4.298	2.016	0.336 (0.122)	0.007
(4) Sensor (Yesterd.)	3.706	1.000	4.547	2.133	1.000 (0.000)	.

Notes: The "Beta's" are from separate regressions for each type of survey question, where the independent variable is the sensor measure for the day before the survey (the benchmark). Standard errors are in parentheses. The last column shows p-values for the same regression.

Table 4
Tests for mean-reverting measurement error - first month.

	(1)	(2)	(3)	(4)	(5)	(6)
	Mean	Ratio to benchmark (Means)	Variance	Ratio to benchmark (Variance)	Beta (SE)	P-Val
(1) Surv (Detail)	3.388	0.877	3.111	0.754	0.056 (0.079)	0.479
(2) Surv (Detail)- Adult	3.193	0.826	1.896	0.460	0.019 (0.054)	0.732
(3) Surv (Aggr.)	3.573	0.925	4.298	1.042	0.361 (0.130)	0.006
(4) Sensor (August)	3.864	1.000	4.125	1.000	1.000 (0.000)	.

Notes: The "Beta's" are from separate regressions for each type of survey question, where the independent variable is the sensor measure for August, the first month of use (the benchmark). Standard errors are in parentheses. The last column shows p-values for the same regression.



Notes: For the sensor data, we classify usage by whether the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot. We then calculated for each sensor the percentage of days that the light was on across all days that the sensor worked and then used this information to calculate the average across all sensors. In the survey, we asked about activities and light use for each time slot separately. Sample is restricted to sensors that worked until the end of the study and households where we have endline data for adults and pupils.

Fig. 5. Use across the day: sensor vs. survey data (adults only).

survey we did not ask the Detailed Questions about use during the daytime and late at night, and hence these slots are, by design, empty. As expected, adults' reported use only reflects a fraction of total use and is lower for each time slot.[16] Fig. 6 compares the combined answers of adults and pupils with sensor data. While the reports of usage during the evening hours seem to match the sensor data well, some children seem to overreport use in the early morning hours.

Comparing the averages of the Aggregated Question and the Detailed (or time diary) Questions, we observe that the Aggregated Question is much closer to the sensor value and statistically indistinguishable from the sensor data (Table 6, column 3). Moreover, the correlation coefficients in Table 5 suggest that at the unit (household) level the Detailed Questions are also less correlated with use measured by the sensors than the Aggregated Question. We do not believe this finding is driven by the fact that the Detailed Questions only ask about nine of the 24 hours of a day. The sensor data show very little solar light usage during daytime hours (see Figs. 5 and 6). This result might be surprising, given that asking individuals about each time slot separately (time diaries) is often considered best practice to measure time use in the literature. It is interesting that this much more lengthy and costly survey method did not correlate more with sensor data than simply asking for a global average of light use for the previous day.

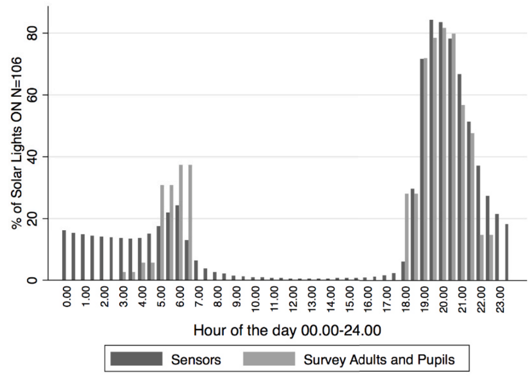
4.4. Precision gains with sensor data

While self-reported daily use of solar lights looks very similar to sensor data on average (Tables 2 and 6), the individual observations are not highly correlated (Table 5). In particular, correlation coefficients of the Detailed Questions are very small, suggesting that the data are very noisy.

An advantage of sensor data is that random measurement error is reduced and precision gains can be achieved, which enables researchers to detect smaller differences in use among sub-groups or to use smaller sample sizes than are necessary when using surveys to measure the impact of a new technology on behavior. For example, in Rom and Günther (2019), we were interested to know whether households that received a free light used it less than households that paid for the light, in order to analyze the potential effectiveness of subsidies in increasing technology adoption. One might expect that the selection of households that purchase a solar light use it more than the average household who gets it for free,[17] as households planning to use the light a lot are more likely to buy one (selection effect); simply having already paid for the solar light may also make households more likely to use it (sunk cost effect). While we cannot differentiate between these two possible effects, we can estimate if both combined lead to different use patterns, which seems relevant for a government considering subsidies, for

[16] This is consistent with adults' answers to a separate question, to which over 70% responded that their school-aged children were the main users of the solar light.

[17] We are fully aware of the fact that there is a selection of people in the purchasing treatment, but still think this is an interesting comparison to make, as pioneered by a paper of Cohen and Dupas (2010). We analyze the correlates of positive purchasing decisions in Appendix B, Table B.8.



Notes: For the sensor data, we classify usage by whether the solar light was used for more than 15 minutes without interruption during the relevant half-hour slot. We then calculated for each sensor the percentage of days that the light was on across all days that the sensor worked and then used this information to calculate the average across all sensors. In the survey, we asked about activities and light use for each time slot separately (in this graph we count the light as being used if either the pupil or the adult indicated that they used it). Sample is restricted to sensors that worked until the end of the study and households where we have endline data for adults and pupils.

Fig. 6. Use across the day: sensor vs. survey data (adults and pupils).

Table 5
Correlations light used (Hrs) per day: Survey and sensor data.

	(1)	(2)	(3)
	Surv (Detail)	Surv (Detail) Adult	Surv (Aggr.)
(1) Sens (Aug)	0.064	0.027	0.338
	(215)	(215)	(161)
(2) Sens (Yest.)	0.065	0.127	0.372
	(147)	(147)	(125)

Notes: Table shows correlations between variables in Rows and Columns. Number of observations are shown in brackets. Number of observations varies since we do not have sensor data for all sensors until the end of the study and we do not have all survey measures for all observations.

Table 6
Differences in light use (Hrs) per day: Survey and sensor data.

	(1)	(2)	(3)
	Surv (Detail)	Surv (Detail) Adult	Surv (Aggr.)
(1) Sens (Aug)	0.450**	0.645***	0.288
	(215)	(215)	(161)
(2) Sens (Yest.)	0.049	0.532***	0.285
	(147)	(147)	(125)

Notes: This table shows differences between variables in Rows (Sensor Data) and Columns (Survey Data). ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Number of observations are shown in brackets. Number of observations varies since we do not have sensor data for all sensors until the end of the study and we do not have all survey measures for all observations.

example. Moreover, we are interested in whether poorer households use solar lights more. Unlike purchasing kerosene for lighting, the marginal cost of an additional hour of solar light is effectively zero. Therefore, we expect more budget-constrained households to use more light once they get access to a solar light. Last, more educated and larger households might use it more and households with previous access to electricity less. For this type of sub-group analysis, precision gains could allow you to detect differences that might not otherwise be evident.

In Table 7, we show how the use of solar lights varies for different sub-groups across both types of usage data. Survey answers include the entire sample of households with a solar light (N = 495), i.e., even for those solar lights that were not equipped with a sensor (see Section 2). Results are not different if we restrict the sample to those households that had a solar light with sensor. Looking at sensor data, a number of variables suggest that poorer households use the solar light more. For example, we find that households that frequently cut meals, households with earth-floor houses, households with a lower wealth index, and households without modern energy access use the solar light more. For the survey data, this correlation is only statistically significant at a 5% level for households that frequently cut meals, even though the sample size is larger (N = 495) than for the sensor sample (N = 220). These results highlight the precision gains possible with sensor data when analyzing correlates of usage. Had we only drawn conclusions from the survey data, we would not have had sufficient evidence to conclude that poorer households use the solar light more, and might have inferred that there is no meaningful relationship between indicators of poverty and solar light usage.

5. Conclusion

There are a number of challenges with self-reported data on technology adoption, including social desirability bias, biases related to the fact that respondents feel observed, and accurate information recall. Sensors can provide more accurate, precise data at a higher frequency than self-reported data. Hence, they can reduce the cost of analyzing behavioral change. In addition, they can help us understand biases and improve survey design, as we can test different survey techniques and compare responses to data collected with sensors. Sensor technology has the potential to transform how we measure human behavior and track the performance of policies and programs, however, there are still challenges to be overcome regarding the functionality of the technology over time. More field testing and training for social science researchers in charge of dealing with these new tools is needed (see Section 2.3 and Appendix C for more details).

While a number of studies have used sensors to measure the adoption of cookstoves (Wilson et al., 2016; Ramanathan et al., 2016; Thomas et al., 2013), this study is the first to use sensors to measure the adoption of solar lights on a large scale. Gandhi et al. (2016) used sensors to measure solar light adoption in only 37 households over less than two weeks. We were able to use sensors to collect information about solar

Table 7
Correlates of light use.

VARIABLES	(1) Sensor (Hrs) First Month	(2) Survey (Hrs) with Sensors	(3) Survey (Hrs) All	(4) Sensor (Hrs) First Month	(5) Survey (Hrs) with Sensors	(6) Survey (Hrs) All	(7) Sensor (Hrs) First Month	(8) Survey (Hrs) with Sensors	(9) Survey (Hrs) All	(10) Sensor (Hrs) First Months	(11) Survey (Hrs) with Sensors	(12) Survey (Hrs) All
Free Light	0.203	−0.246	−0.120									
	(0.348)	(0.362)	(0.175)									
Earth Floor				1.065***	0.364	0.354*						
				(0.325)	(0.462)	(0.202)						
Freq Cut Meal							0.246*	0.390**	0.243***			
							(0.136)	(0.162)	(0.087)			
Wealth Index										−0.177*	0.024	0.153
										(0.100)	(0.114)	(0.099)
Constant	3.712***	3.756***	3.505***	2.946***	3.261***	3.133***	3.685***	3.297***	3.250***	4.764***	3.394***	2.647***
	(0.314)	(0.306)	(0.134)	(0.288)	(0.426)	(0.178)	(0.160)	(0.185)	(0.102)	(0.508)	(0.545)	(0.466)
Observations	220	161	495	215	161	495	215	161	495	164	120	293
R-squared	0.002	0.003	0.001	0.038	0.004	0.004	0.016	0.041	0.020	0.019	0.000	0.018
Mean	3.864	3.566	3.434	3.864	3.566	3.434	3.864	3.566	3.434	3.864	3.566	3.434

Notes: Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Wealth index has fewer observations since we only collected data on assets for a subgroup. The wealth index includes information about the household's ownership of bikes, motorbikes, cars, stoves, radios, wall clocks, tin lamps, kerosene lanterns, solar lanterns, electrical lanterns, tables, beds, bed nets, chairs, sofa pieces, jembes, car batteries, animal charts, horses, cattle, goats, sheeps, chickens, pigs, mobile phones, and sim cards.

light use in over 200 households, some of which purchased the solar light, while others received one for free.

We find that households use solar lights for around 4 hours per day on average and that fewer than 5% of households use the solar lights infrequently. Adoption of solar lights is much higher than what recent studies on cookstove adoption have found (Wilson et al., 2016). We also used sensor data to test what types of survey questions led to more accurate answers and whether differences between self-reported information and sensor data were particularly large for certain sub-groups.

A number of results seem especially relevant: first, as opposed to a number of papers on cookstoves (Wilson et al., 2016; Ramanathan et al., 2016; Thomas et al., 2013), and the small-scale study on solar lights (Gandhi et al., 2016), we do not find that households systematically overreport use of solar lights. However, in line with the findings of these studies, overreporting was more likely when the household used the solar light very little, which could be explained by social desirability bias. In addition, and consistent with mean-reverting measurement error (Bound et al., 2001; Gibson et al., 2015), we find that households that use the solar light a lot tend to underreport use, which, to our knowledge, has not been found before. As adoption of solar lights was nearly universal, we do not find evidence for systematic overreporting on average. In addition to the difference in adoption rates between cookstoves and solar lights, the nature of solar light usage is also very different. Solar lights can be used by many household members throughout the entire day and in ways that are not visible to the respondent, whereas the use of cookstoves is typically reserved for a few household members and for a limited number of times at specific times of day. These differences might explain why underreporting was more common in our case.

Second, while the reported hours of use per day are quite similar on average, answers from individual households correlated very little with the information we got from the sensors, suggesting that random errors are very large in survey data on technology use.

Third, we find that asking aggregated questions about use provides more accurate information than asking for each time slot separately (time diary). This result is surprising, given that time diaries are considered the gold standard in time-use data collection. However, there are still very few papers confirming the validity of this claim in developing countries (Seymour et al., 2017). The lack of correlation between the time diary survey responses and the sensor data could also be due to survey design issues, as we did not ask for every time slot throughout the day and we did not survey every household member.

Finally, we find that, as predicted by the Hawthorne effect, more frequent visits from surveyors in the beginning of the study did increase use initially. This difference disappeared once the visits stopped. Wilson et al. (2016) made a similar discovery when studying cookstoves.

We are not suggesting that sensors should replace surveys or that they should or can be used in every study of technology adoption. Many questions about adoption, and the use and impact of new technologies cannot be answered with sensors alone. In addition, sensors still require careful and time-intensive field testing, as frequent failures still pose challenges in many studies, including ours (Wilson et al., 2016; Piedrahita et al., 2016). Our results, however, highlight how sensors can provide more accurate and precise information, which seems particularly relevant when social desirability is expected to be high. While it might be too early to draw general conclusions, a number of studies, including ours, suggest that the overreporting of use is mostly a problem when adoption is low, and hence that it is particularly important to receive an objective measurement in such cases. We also observe that while survey and sensor measurements were similar on average, they did not agree for individual households. Hence, sensors might be particularly relevant when researchers want to conduct sub-group analyses or use smaller sample sizes.

Finally, sensor data can help us better understand how to improve study and survey design, since they provide a credible benchmark to test different types of survey questions and interactions between surveyors and respondents.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

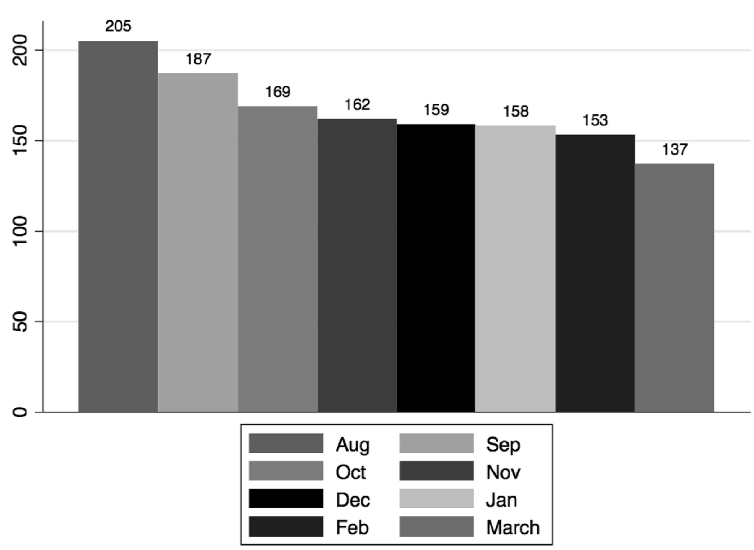
We thank Kat Harrison, Professor Edward Miguel, Professor Lorenzo Casaburi, Dr. Nick Lam, the seminar participants at the Development Economics seminar at the University of California, Berkeley, the CSAE Conference, the PEGNet Conference, and the CEPE/CER-ETH Lunch Seminar Series, and two anonymous reviewers for useful comments. We also thank Carol Nekesa, Lisa Schauss, Charles Amuku, Selina Obwora, Erick Bwire, and the entire field team from Innovations for Poverty Action Kenya for outstanding field work. This research was made possible with the support and cooperation of Acumen, Bonsai Systems, Innovations for Poverty Action, SolarAid, and REMIT Kenya Consulting. This work was supported by Google.org and the Development Economics Group and the NADEL Center for Development and Cooperation at ETH Zurich.

Appendix

A. Figures

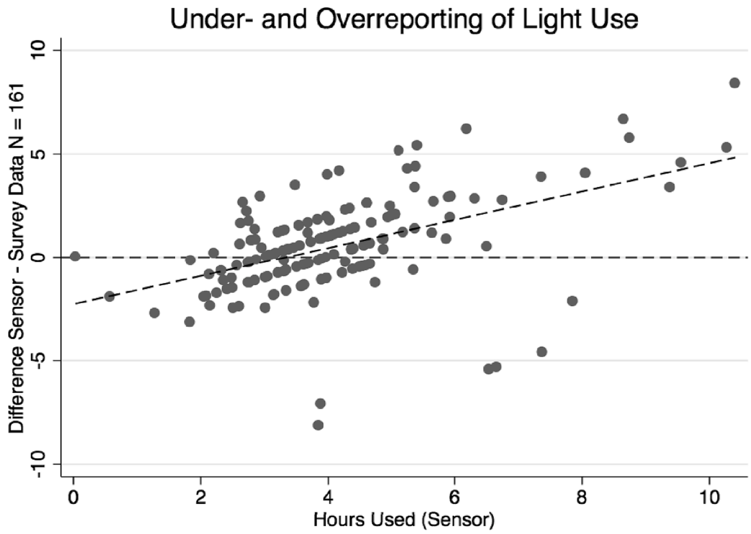
Fig. A.1. Sun King Eco solar light.

Fig. A.2. Sun King Mobile solar light.



Notes: This graph shows the number of sensors that worked until the end of the indicated month.

Fig. A.3. Number of working sensors by the end of each month.



Notes: This graph shows the correlation between the difference of sensor data and the survey data (Aggregated Question) and average hours used per day according to the sensor data. Positive values on the y axis indicate that respondents underreported use, while negative values suggest that they overreported use.

Fig. A.4. Under- and overreporting of use.

B. Tables

Table B.1
Summary statistics.

Variable	(1) Mean	(2) SD	(3) N
Earth Floor	0.841	0.366	214
Freq Cut Meal	0.623	1.024	215
Wealth Index	4.789	1.479	164
HH Head Yrs of Schooling	6.418	3.747	208
HH Size	6.559	2.126	220
Energy Access	0.065	0.247	215

Notes: Wealth index has fewer observations since we only collected data on assets for a subgroup.

Table B.2
Use previous month as predictor for survival.

VARIABLES	(1) Sensor Hrs- Aug	(2) Sensor Hrs- Sept	(3) Sensor Hrs- Oct	(4) Sensor Hrs- Nov	(5) Sensor Hrs- Dec	(6) Sensor Hrs_ Jan	(7) Sensor Hrs- Feb
Stopped Working in Sept	0.819						
	(0.538)						
Stopped Working in Oct		1.263**					
		(0.507)					
Stopped Working in Nov			−0.396				
			(0.268)				
Stopped Working in Dec				−2.237**			
				(0.971)			
Stopped Working in Jan					0.460***		
					(0.150)		
Stopped Working in Feb						−2.046**	
						(0.803)	
Stopped Working in March							0.031
							(0.482)

(continued on next page)

Table B.2 (*continued*)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Sensor Hrs- Aug	Sensor Hrs- Sept	Sensor Hrs- Oct	Sensor Hrs- Nov	Sensor Hrs- Dec	Sensor Hrs_ Jan	Sensor Hrs- Feb
Constant	3.759***	3.828***	4.077***	4.095***	3.964***	4.041***	4.097***
	(0.148)	(0.160)	(0.146)	(0.161)	(0.150)	(0.143)	(0.154)
Observations	205	187	169	162	159	158	153
R-squared	0.013	0.031	0.002	0.022	0.000	0.040	0.000
N droped next Month	18	18	7	3	1	5	16

Notes: Robust standard errors in parentheses.***p < 0.01, **p < 0.05, *p < 0.1.

Table B.3
Correlates of sensor attrition before the end of the study.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Attrition End	Attrition End	Attrition End	Attrition End	Attrition End	Attrition End
Free Light	0.104					
	(0.075)					
Earth Floor		0.063				
		(0.084)				
Freq Cut Meal			0.034			
			(0.032)			
HH Head Yrs of Schooling				0.001		
				(0.009)		
HH Size					0.035**	
					(0.015)	
Energy Access						−0.033
						(0.126)
Constant	0.607***	0.265***	0.295***	0.298***	0.088	0.318***
	(0.066)	(0.076)	(0.037)	(0.064)	(0.097)	(0.033)
Observations	215	214	215	204	215	215
R-squared	0.010	0.002	0.006	0.000	0.026	0.000
Mean	0.316	0.316	0.316	0.316	0.316	0.316

Notes: Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Attrition is a dummy = 1 if the sensor stopped working before the last month of the study and = 0 if it worked until the end of the study.

Table B.4
Correlates of sensor attrition before the end of the first month (Aug).

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	Attrition Aug	Attrition Aug	Attrition Aug	Attrition Aug	Attrition Aug	Attrition Aug
Free Light	0.020					
	(0.036)					
Earth Floor		0.013				
		(0.045)				
Freq Cut Meal			0.012			
			(0.020)			
HH Head Yrs of Schooling				0.005		
				(0.005)		
HH Size					0.008	
					(0.010)	
Energy Access						−0.075***
						(0.019)
Constant	0.054*	0.059	0.062***	0.038	0.018	0.075***
	(0.030)	(0.041)	(0.020)	(0.034)	(0.064)	(0.019)
Observations	220	214	215	208	220	215
R-squared	0.001	0.000	0.002	0.005	0.004	0.005
Mean	0.0680	0.0680	0.0680	0.0680	0.0680	0.0680

Notes: Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Attrition is a dummy = 1 if the sensor stopped working before the last month of the study and = 0 if it worked until the end of the study.

Table B.5
Use across months.

VARIABLES	(1)
	Sensor Hrs
September	0.183
	(0.254)
October	0.387
	(0.239)
November	0.384
	(0.250)
December	0.246
	(0.240)
January	0.242
	(0.233)
February	0.302
	(0.233)
March	0.325
	(0.230)
Observations	1,096
R-squared	0.004
Mean Sensor	4.053

Notes: Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Left out group is August. We first calculated the average use per month for the 137 sensors we have data for until the end of March (rather than the end of the study, which was in mid-March). Mean use is across all months. Total observations are months (8) x number of sensors (137).

Table B.6
Use across weekdays.

VARIABLES	(1)
	Sensor Hrs
Tuesday	0.062*
	(0.032)
Wednesday	0.016
	(0.029)
Thursday	0.016
	(0.030)
Friday	−0.008
	(0.034)
Saturday	−0.096**
	(0.046)
Sunday	−0.174***
	(0.039)
Observations	959
R-squared	0.002
Mean Use	4.022

Notes: Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$. Left out group is Monday. We first calculated the average use per weekday for the 137 sensors we have data for until the end of March (rather than the end of the study, which was in mid-March). Mean use is across all weekdays. Total observations are number of days (7) x number of sensors (137).

Table B.7
Correlates of differences between sensor and survey estimates.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)
Additional Visits	0.588							
	(0.422)							
Free Solar Light		0.224						
		(0.455)						
Earth Floor			−0.242					
			(0.503)					
Freq Cut Meal				−0.163				
				(0.263)				
Wealth Index					0.077			
					(0.170)			

(*continued on next page*)

Table B.7 (*continued*)

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)	Diff Sens (Yes)-Surv (Aggr)
HH Head Yrs of Schooling						0.009		
						(0.056)		
HH Size							0.102	
							(0.086)	
Energy Access								1.696**
								(0.816)
Constant	0.526**	0.572	0.949**	0.835***	0.423	0.740*	0.095	0.627***
	(0.253)	(0.387)	(0.448)	(0.232)	(0.833)	(0.422)	(0.588)	(0.208)
Observations	146	146	145	146	112	141	146	146
R-squared	0.013	0.001	0.001	0.004	0.002	0.000	0.008	0.031
Mean	0.744	0.744	0.744	0.744	0.744	0.744	0.744	0.744

Notes: Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1. Includes all 146 sensors for which we have data for the day before endline data collection as well as the aggregated survey measure. Column 4 has fewer observations since we only collected data on assets for a sub-group.

Table B.8
Correlates of purchasing decision.

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Bought Solar	Bought Solar	Bought Solar	Bought Solar	Bought Solar	Bought Solar	Bought Solar
Price 400	0.393***	0.390***	0.393***	0.360***	0.402***	0.395***	0.393***
	(0.046)	(0.046)	(0.046)	(0.080)	(0.045)	(0.046)	(0.046)
Price 700	0.080*	0.078	0.080*	0.061	0.082*	0.081*	0.080*
	(0.048)	(0.048)	(0.048)	(0.079)	(0.047)	(0.048)	(0.048)
Earth Floor		−0.127**					
		(0.064)					
Iron Roof			0.004				
			(0.039)				
Freq. cut Meal				−0.042			
				(0.044)			
HH Head Yrs of schooling					0.019***		
					(0.005)		
HH Size						0.008	
						(0.009)	
Electricity Access							0.015
							(0.130)
Constant	0.294***	0.410***	0.292***	0.357***	0.171***	0.241***	0.294***
	(0.033)	(0.067)	(0.041)	(0.115)	(0.046)	(0.068)	(0.033)
Observations	600	596	600	204	599	600	600
R-squared	0.118	0.121	0.118	0.105	0.139	0.119	0.118
Mean	0.457	0.457	0.457	0.457	0.457	0.457	0.457

Notes: Robust standard errors in parentheses. ***p < 0.01, **p < 0.05, *p < 0.1.

C. *Lessons learned from using sensors to study technology adoption in low-income settings*

First, it is critical to thoroughly pre-test sensor technology (both the sensor and the application to access the data) at a reasonably large scale in the field and to only roll out the study once all problems are solved. Often, engineering teams designing sensors are used to small sample sizes where technological challenges can be fixed along the way. It might make sense to agree in advance on a threshold of acceptable failure rates in the pilot as a commitment device. For example, we installed the sensors in a pre-existing product that was not designed to hold a sensor, thus, several sensors probably stopped working due to an imperfectly soldered connection between the sensor and the existing hardware, which also led to more light breakages. An additional challenge we had was that the application designed to access the data from the sensors initially did not work reliably and it took us time to determine the extent of the problem. In the meantime, our field officers had to return to the same households multiple times to ensure the data were collected. Since some of the sensors stopped working before the application was fully functioning, we lost a significant amount of data. Such issues could possibly be avoided by testing the sensors and associated technology extensively in the field and under a variety of realistic circumstances to determine vulnerabilities to contextual factors that are hard to recreate in the lab.

Second, if the sensor is not constantly transmitting data to a central storage location throughout the study, we recommend doing a first round of sensor data collection immediately after installation and distribution (i.e., immediately after baseline) to guard against challenges linked to sensor attrition, which turned out to be a major problem in our study. Collecting data early not only ensures some data is collected from the maximum number of sensors, but can also help identify problems before they become widespread.

As a result of the two issues mentioned above, our third recommendation is to create a very detailed protocol on how to proceed if a sensor or the host technology stops working and, ideally, to include it in the pre-analysis plan. Both sensors and solar lights stopped working more often than we expected, and it was not possible to distinguish from the sensor data if the solar light broke because of the sensor or vice versa. It is therefore important to remember that both human error and technology failure are possible when building up a testing protocol. We suggest developing clear instructions about what to do if the analyzed technology or the sensor fails and to keep detailed information about replacements in order to easily account for these sensors in the analysis. Furthermore, we recommend allocating a sufficient amount of staff time to this effort. In cases where the sensor technology has not been tested extensively in the field over long periods of time, we also recommend designing the research in such a way that the most important questions can be answered even if there is a lot of sensor attrition.

Our final recommendation is to take time to explain the sensor technology to partner organizations and the community. For example, we co-wrote a letter with the engineering team that developed the sensors explaining the functionality of the sensors to our partner organization. We also tested the acceptability of the sensors with a separate sample and developed a detailed script to explain the sensors to users. This script was written with guidance from our local partners, who are very familiar with the resident community. In addition, we provided respondents with our contact information in case of problems. We had no problems with regard to the acceptability of the sensors in the local community, but we imagine that this is highly context dependent.

D. Survey questions

Aggregated Question

- Do you own one or several lanterns? Options: yes/no
  - If yes: Does any of your solar lanterns still work? Options: yes/no
  - If yes: Yesterday, for how many hours did you use a solar lantern?

Options: 0 h–24 h

Time Diary Questions

- What did you do between XX:XX and XX:XX?

  Options:
  same as in previous time slot,
  at work (non-agricultural work)
  barber
  salon
  bathe
  dress
  brewing alcohol
  care for children/sick/elderly
  clean
  dust, sweep
  wash dishes or clothes
  ironing
  other household chores cook
  prepare food
  discuss activities of the next day with partner
  doctor/hospital
  visit
  eat
  farm work
  fetch water
  firewood
  fishing or hunting
  funeral/wedding activities
  help homework
  herding animals/work with livestock
  listen to radio
  other religious activity (e.g., study, group)
  participate in community activities/meetings/voluntary work
  play sports
  pray
  prepare children for school
  read book
  repairs around/on home
  rest
  sewing/fixing
  clothes shop for family
  sleep
  socialize with other household members
  socialize with people outside of the household
  spend time with spouse/partner
  study/attend class
  travel by bicycle
  travel by foot
  travel by motorized means

visit/entertain friends
watch TV
Other

- What lighting source did you use for this activity, if any?
  Options:
  Electricity powered lighting
  Solar home system powered lighting
  Tin Lamp
  Kerosene lantern/Hurricane
  Fire Wood
  Battery powered torch/lantern
  Candle Solar lantern/solar torch
  Pressurized Kerosene Lantern
  Other rechargeable lantern
  Cell phone light
  No lighting used
  Matchsticks
  Other

References

Adair, J.G., Sharpe, D., Huynh, C.L., 1989. Hawthorne control procedures in educational experiments: a reconsideration of their use and effectiveness. Rev. Educ. Res. 59 (2), 215–228.
Angrist, J., Pischke, J., 2010. The credibility revolution in empirical economics: how better research design is taking the con out of econometrics. J. Econ. Perspect. 24 (2), 3–30.
Arthi, V.S., Beegle, K., De Weerdt, J., Palacios-Lopez, A., 2016. Not Your Average Job: Measuring Farm Labor in Tanzania. Policy Research Working Paper No. 7773. World Bank Group, Washington, D.C.
Banerjee, A.V., Duflo, E., Glennerster, R., Kothari, D., 2010. Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives. BMJ 340, c2220.
Bardasi, E., Beegle, K., Dillon, A., Serneels, P., 2011. Do labor statistics depend on how and to whom the questions are asked? Results from a survey experiment in Tanzania. World Bank Econ. Rev. 25 (3), 418–447.
Beaman, L., Magruder, J., Robinson, J., 2014. Minding small change among small firms in Kenya. J. Dev. Econ. 108, 69–86.
Beegle, K., Carletto, C., Himelein, K., 2012. Reliability of recall in agricultural data. J. Dev. Econ. 98 (1), p34–41.
Bertrand, M., Mullainathan, S., 2001. Do people mean what they say? Implications for subjective survey data. Am. Econ. Rev. 91 (2), p67–72.
Bonggeun, K., Solon, G., 2005. Implications of mean-reverting measurement error for longitudinal studies of wages and employment. Rev. Econ. Stat. 87 (1), p193–196.
Bound, J., Krueger, A.B., 1991. The extent of measurement error in longitudinal earnings data: do two wrongs make a right? J. Labor Econ. 9 (1), p1–24.
Bound, J., Brown, C., Duncan, G.J., Rodgers, W.L., 1994. Evidence on the validity of cross-sectional and longitudinal labor market data. J. Labor Econ. 12 (3), p345–368.
Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: Heckman, J.J., Leamer, E.E. (Eds.), Handbook of Econometrics, vol. 5. Elsevier, Amsterdam, pp. 3705–3843.
Clasen, T., Fabini, D., Boisson, S., Taneja, J., Song, J., Aichinger, E., Bui, A., Dadashi, S., Schmidt, W.P., Burt, Z., Nelson, K.L., 2012. Making sanitation count: developing and testing a device for assessing latrine use in low-income settings. Environ. Sci. Technol. 46 (6), p3295–3303.
Cohen, J., Dupas, P., 2010. Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. Q. J. Econ. 125 (1), p1–45.
Daniels, L., 2001. Testing alternative measures of microenterprise profits and net worth. J. Int. Dev.: J. Dev. Assoc. 13 (5), p599–614.
Das, J., Hammer, J., Sanchez-Paramo, C., 2012. The impact of recall periods on reported morbidity and health seeking behavior. J. Dev. Econ. 98, p76–88.
De Mel, S., McKenzie, D.J., Woodruff, C., 2009. Measuring microenterprise profits: must we ask how the sausage is made? J. Dev. Econ. 88 (1), 19–31.
Deaton, A., Grosh, M., 2000. Consumption. In: Grosh, M., Glewwe, P. (Eds.), Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of the Living Standards Measurement Study. World Bank, Washington, D.C., pp. 91–133
Duflo, E., Glennerster, R., Kremer, M., 2008. Using randomization in development economics research: a toolkit. In: Schultz, T., Strauss, J. (Eds.), Handbook of Development Economics, vol. 4. Elsevier, Amsterdam, pp. 3895–3962. https://econpapers.repec.org/bookchap/eeedevchp/5-61.htm.
Gandhi, A., Frey, D., Lesniewski, V., 2016. Assessing solar lantern usage in Uganda through qualitative and sensor-based methods. In: 2016 IEEE Global Humanitarian Technology Conference. IEEE, New Jersey. Seattle, 13-16 October.
Garn, J.V., Sclar, G.D., Freeman, M.C., Penakalapati, G., Alexander, K.T., Brooks, P., Rehfuess, E., Boisson, S., Medlicott, O., Clasen, T.F., 2017. The impact of sanitation interventions on latrine coverage and latrine use: a systematic review and meta-analysis. Int. J. Hyg Environ. Health 220 (2/Part B), p329–340.
Gautam, S., 2017. Quantifying Welfare Effects in the Presence of Externalities: an Ex-Ante Evaluation of a Sanitation Intervention. Working paper.
Gibson, J., Beegle, K., De Weerdt, J., Friedman, J.A., 2015. What does variation in survey design reveal about the nature of measurement errors in household consumption? Oxf. Bull. Econ. Stat. 77 (3), p466–474.
Greenlight Planet, 2016. Product Sheet Sun king Eco [Online]. Available at: http://www.greenlightplanet.com. (Accessed 13 October 2017).
Grosh, M., Glewwe, P., 2000. Designing Household Survey Questionnaires for Developing Countries. World Bank Publications, 25338. The World Bank, Washington DC.
Hanna, R., Duflo, E., Greenstone, M., 2016. Up in smoke: the influence of household behavior on the long run impact of improved cooking stoves. Am. Econ. J. Econ. Pol. 8 (1), p80–114.
Innovations for Poverty Action, 2016. Sensing Impacts: Remote Monitoring Using Sensors [Online]. Available at: https://www.poverty-action.org/sites/default/files/publications/Goldilocks-Deep-Dive-Sensing-Impacts-Remote-Monitoring-using-Sensors_4.pdf. (Accessed 12 August 2018).
Kane, T.J., Rouse, C.E., Stagier, D., 1999. Estimating Returns to Schooling when Schooling Is Misreported. National Bureau of Economic Research. Working Paper 7235. Available at: http://www.nber.org/papers/w7235. (Accessed 20 April 2020).
Leonard, K., Masatu, M.C., 2006. Outpatient process quality evaluation and the Hawthorne Effect. Soc. Sci. Med. 63 (9), p2330–2340.
Levitt, S.D., List, J.A., 2011. Was there really a Hawthorne effect at the Hawthorne plant? An analysis of the original illumination experiments. Am. Econ. J. Appl. Econ. 3 (1), p224–p238.
Lighting Global, 2015. Product Verification Sheet [online]. Available at: http://www.lightingglobal.org/products/glp-sunkingeco. (Accessed 12 September 2017).
Loken, E., Gelman, A., 2017. Measurement error and the replication crisis. Science 355 (6325), p584–585.
Mathiowetz, N.A., Groves, R.M., 1985. The effects of respondent rules on health survey reports. Am. J. Publ. Health 75 (6), 639–644.
McCambridge, J., Witton, J., Elbourne, D.R., 2014. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. J. Clin. Epidemiol. 67 (3), p267–277.
Mills, E., 2003. Technical and economic performance analysis of kerosene lamps and alternative approaches to illumination. In: Developing Countries. Available at: https://www.worleyparsons.com.
O'Neill, D., Sweetman, O., 2012. The Consequences of Measurement Error when Estimating the Impact of BMI on Labour Market Outcomes. IZA Discussion Paper No. 7008. [Online]. Available at: https://ssrn.com/abstract=2177206.
Piedrahita, R., Dickinson, K.L., Kanyomse, E., Coffey, E., Alirigia, R., Hagar, Y., Rivera, I., Oduro, A., Dukic, V., Wiedinmyer, C., Hannigan, M., 2016. Assessment of cookstove stacking in Northern Ghana using surveys and stove use monitors. Energy Sustain. Dev. 34, p67–76.
Pillarisetti, A., Allen, T., Ruiz-Mercado, I., Edwards, R., Chowdhury, Z., Garland, C., Hill, L.D., Johnson, M., Litton, C.D., Lam, N.L., Pennise, D., Smith, K.R., 2017. Small, smart, fast, and cheap: microchip-based sensors to estimate air pollution exposures in rural households. Sensors 17 (8), 1879.
Pischke, J., 1995. Measurement error and earnings dynamics: some estimates from the PSID validation study. J. Bus. Econ. Stat. 13 (3), p305–314.
Ramanathan, T., Ramanathan, N., Mohanty, J., Rehman, I., Graham, E., Ramanathan, V., 2016. Wireless sensors linked to climate financing for globally affordable clean cooking. Nat. Clim. Change 7, p44–47.
Rom, A., Günther, I., 2019. Decreasing Emissions by Increasing Energy Access? Evidence from a Randomized Field Experiment on Off-Grid Solar. ETH Working Paper. [Online] Available at: https://ethz.ch/content/dam/ethz/special-interest/gess/nadel-dam/documents/2019.08.21_Emissions_Access.pdf.
Ruiz-Mercado, I., Canuz, E., Walker, J.L., Smith, K.R., 2013. Quantitative metrics of stove adoption using Stove Use Monitors (SUMs). Biomass Bioenergy 57, p136–148.

Serneels, P.M., Beegle, K.G., Dillon, A.S., 2016. Do returns to education depend on how and whom you ask? Econ. Educ. Rev. 60, p5–19.

Seymour, G., Malapit, H.J., Quisumbing, A.R., 2017. Measuring Time Use in Development Settings. Policy Research Working Paper No. 8147. The World Bank, Washington D.C.

Smits, J., Günther, I., 2018. Do financial diaries affect financial outcomes? Evidence from a randomized experiment in Uganda. Dev. Eng. 3, p72–82.

Thomas, E.A., Barstow, C.K., Rosa, G., Majorin, F., Clasen, T., 2013. Use of remotely reporting electronic sensors for assessing use of water filters and cookstoves in Rwanda. Environ. Sci. Technol. 47, p13602–13610.

Wilson, D., Coyle, J., Kirk, A., Rosa, J., Abbas, O., Adam, M., Gadgil, A., 2016. Measuring and increasing adoption rates of cookstoves in a humanitarian crisis. Environ. Sci. Technol. 50 (15), p8393–8399.

Zwane, A., Zinman, J., Van Dusen, E., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D., Hornbeck, R., Gine, X., Duflo, E., Devoto, F., Crepon, B., Banerjee, A., 2011. Being surveyed can change later behavior and related parameter estimates. Proc. Natl. Acad. Sci. 108 (5), p1821–1826.