

“Call Me Educated: Evidence from a Mobile Monitoring Experiment in Niger”

Jenny C. Aker and Christopher Ksoll *

January 2017

Abstract. In remote rural areas of developing countries, teacher absenteeism is a widespread problem. We report the results of a randomized evaluation of an adult education program in Niger, in which a subset of villages were randomly assigned to a mobile phone monitoring intervention, without a financial incentive. Students in villages with the mobile monitoring intervention achieved test scores that were .07-.18 standard deviations higher as compared to those in standard adult education villages. These results suggest that mobile phone technology can be used as a monitoring tool to improve learning outcomes.

JEL codes: D1, I2, O1, O3

Keywords: Adult education, teacher absence, monitoring, information technology, Niger

*Jenny C. Aker, The Fletcher School and Department of Economics, Tufts University, 160 Packard Avenue, Medford, MA 02155; Jenny.Aker@tufts.edu. Christopher Ksoll, Mathematica Policy Research, Oakland, CA; christopher.ksoll@gmail.com. We thank Michael Klein, Julie Schaffner, Shinsuke Tanaka and seminar participants at Tufts University, the Center for Global Development, IFPRI, the World Bank, University of Washington, Georgetown University, and the University of South Carolina for helpful comments. We are extremely grateful for funding from the DFID Economic and Social Research Council (Grant Number ES/L005433/1).

In rural areas of developing countries, public sector absenteeism – of teachers, doctors, nurses or agricultural extension agents – is a widespread problem. In West Africa, teacher absenteeism is estimated between 27-40% (Transparency International 2013). One potential solution has been to strengthen the monitoring of teachers (Banerjee and Duflo, 2006; Glewwe and Kremer, 2006). Yet despite numerous interventions, such as community-based monitoring, hiring teachers from the community and audits, teacher monitoring continues to be a significant challenge. This is particularly the case in countries with limited infrastructure and weak institutions, where the costs of monitoring are particularly high.

The introduction of mobile phone technology throughout sub-Saharan Africa has the potential to reduce the costs associated with monitoring public employees, such as teachers. By allowing governments and organizations to communicate with remote villages on a regular basis, “mobile monitoring” has the potential to increase the observability of the agents’ effort. Similarly, the reductions in communication costs associated with mobile phone technology could potentially increase community engagement in the monitoring process, thereby providing the community with additional bargaining power.

We report the results of a randomized monitoring intervention in Niger, where a mobile phone monitoring component was added to an adult education program. Implemented in 134 villages in two rural regions of Niger, students followed a basic adult education curriculum, but a subset of the villages also received a monitoring component – weekly phone calls to the teacher, the village chief and two students. No other incentives or formal sanctions were provided in the short-term.

Overall, our results provide evidence that the mobile phone monitoring of teachers substantially improved learning outcomes. While adults in the standard adult education program increased their reading and math test scores by .13-.30 s.d. as compared with the pure control group (who received no adult education program), the mobile monitoring intervention led to an additional increase in test scores, ranging from .07-.18 s.d. The impacts were similar by region and gender. While the impacts

were stronger during the first year for reading, they were stronger in the second year for more complex math tasks. These effects are associated with increased teacher and student motivation, although we do not find any statistically significant effects on teacher absenteeism.

Our finding that monitoring leads to an improvement in skills acquisition contributes to a debate on the effectiveness of teacher monitoring in other contexts (Guerrero et al 2013). Using monitoring and financial incentives in a randomized experiment in India, Duflo, Hanna and Ryan (2012) find that teacher absenteeism fell by 21 percentage points and children's test scores increased by 0.17 s.d. Using a nationally representative dataset of schools in India, Muralidharan et al (2014) find that increased school monitoring is strongly correlated with lower teacher absence, but do not measure effects on learning. Using a matched design in Peru, Cueto et al (2008) find that a program of monitoring and financial incentives increased teacher attendance. And finally, using mobile phone monitoring linked to financial incentives, Cilliers et al (2014) find that local monitoring improves teacher attendance, but primarily when there are financial incentives.¹ Our experiment is somewhat unique in that it did not provide any explicit financial incentives for teachers, which may be easier for governments to implement but less effective in increasing teacher effort.²

The remainder of the paper is organized as follows. Section II provides background on the setting of the research and the research design, whereas Section III presents the model. Section IV describes the different datasets and estimation strategy, and Section V presents the results. Section VI addresses the potential mechanisms and Section VII discusses alternative explanations. Section VIII discusses cost-benefit analyses and Section IX concludes.

II. Research Setting and Experimental Design

¹ Most recently, de Ree et al (2016) estimate the impact of an unconditional doubling of teachers' salaries in Indonesia, finding an improvement in teachers' job satisfaction but no impact on teacher effort or students' learning outcomes.

² Our paper also contributes to the literature on community-based monitoring and inspection systems (Svensson 2007, Olken 2007, Bengtsson and Engstrom 2014).

With a gross national income per capita of \$641, Niger is one of the lowest-ranked countries on the UN's Human Development (UNDP 2014). The country has some of the lowest educational indicators in sub-Saharan Africa, with estimated literacy rates of 15 percent in 2012 (World Bank 2015). Illiteracy is particularly striking among women and within our study region: It is estimated that only 10 percent of women attended any school in the Maradi and Zinder regions.

A. Adult Education and Mobile Monitoring Interventions

Over a two-year period (2014 and 2015), an international non-governmental organization (NGO), Catholic Relief Services, implemented an adult education program in two rural regions of Niger. The intervention was designed to provide five months of literacy and numeracy instruction over two years to illiterate adults, with a total of 10 months of instruction. Courses were held between February and June, with a break between July and January due to the agricultural planting and harvesting season. All classes taught basic literacy and numeracy skills in the native language of the village (Hausa), as well as functional topics on health, nutrition and agriculture. Each village was allocated 50 students for the adult education program, with spots for 35 women and 15 men.³ These fifty students were taught in two literacy classes, separated by gender. Both classes were held five days per week for three hours per day, and were taught by community teachers who were selected and trained in the adult education methodology by the Ministry of Non-Formal Education.⁴ Since men's and women's classes differed by both gender and class size, we are unable to disentangle the differential effects of gender on learning outcomes.

The mobile monitoring component was implemented in a subset of the adult education villages. For this intervention, data collection agents made weekly phone calls to four individuals over a six-week period, calling the literacy teacher, the village chief and two randomly selected students (one

³ This breakdown differs from our previous study, whereby the 50 student slots were equally allocated between men and women (Aker et al 2012). However, the donor for the program wanted to increase women's access to the adult education program, and thereby allocated more slots to women in each village.

⁴ Unlike previous adult education programs in Niger, the same teacher taught both classes in the village. In addition, the differences in class size by gender makes it difficult for us to disentangle the learning effects by gender as compared with differences in the class size.

female and one male). Neither the students nor the teachers were informed about which students would be called each week. No phones were provided to either teachers or students, and mobile phone ownership was not a necessary condition to be chosen for the intervention.⁵ During the phone calls, the field agents used a script to introduce themselves and then asked a series of questions: if the class was held in the previous week, including the number of days the class was held and the number of hours per day; the number of students who attended the class; and if the respondent had any additional information to share. The mobile monitoring component was introduced two months after the start of the adult education program, and students, teachers, nor CRS field staff were not informed of which villages were selected for mobile monitoring prior to the intervention. Thus, CRS and the Ministry conducted its normal (in person) monitoring activities in the region.⁶ In order to better understand how monitoring may have affect teachers' effort, the experimental design was modified slightly during the second year, with a subset of monitoring villages calling teachers only and the remaining villages receiving the full intervention.

While general information on the results of the monitoring calls were shared with CRS on a weekly basis, due to funding constraints, neither CRS nor the Ministry were able to conduct additional follow-up monitoring visits beyond what they had previously planned for the year. In fact, the overall number of monitoring visits was extremely low for all villages over the two-year period. As a result, teachers were not formally sanctioned for less than contracted effort during the adult education classes; rather, teachers only learned whether they would be retained for the second year well after the end of classes in the first year in December. In all, 23 percent of teachers were replaced between the first and

⁵Phone numbers for the students, village chiefs and teachers were obtained during the initial registration phase for the program. If these individuals did not own a mobile phone, we asked for the number of a friend or family member in the village. If the teacher, village chief or students chosen to receive the monitoring call did not own a phone, the number of a friend or family member was called first, and we asked to speak to the selected individual. For the first year, the same two students were called over the six-week period.

⁶In-person monitoring usually involved short visits to the village at any time to see if the adult education materials were in place. Such visits did not necessarily occur during the time of the adult education class, and did not require observing the class or teaching.

second years, although there was no correlation between firing and the mobile monitoring intervention (Table 8).⁷

B. Experimental Design

In 2013, CRS identified over 500 intervention villages across two regions of Niger, Maradi and Zinder. Of these, we stratified by geographic region and sub-region and randomly selected 134 villages to participate in the research. Among these 134 villages, we stratified by regional and sub-regional administrative divisions before randomly assigning villages to the adult education program (starting classes in 2014) or a control group (to start classes in 2016).⁸ Among the adult education villages, villages were then assigned to either the monitoring or no monitoring intervention. In all, 114 villages were assigned to the adult education program and 20 villages were assigned to the pure control group (no adult education classes).⁹ Among the adult education villages, villages were assigned to the “mobile monitoring” and or the “no mobile monitoring” condition. The final sample in this paper is 131 villages, 20 in the pure control group and 111 in the adult education program.¹⁰ A timeline of the implementation and data collection activities is provided in Figure 1.

Within each village, and prior to the baseline, CRS identified eligible students in both the adult education and control villages, for a total of fifty students per village. While this was intended to be 35 women and 15 men per village, in some villages, only women were registered for the program.

⁷ While CRS did have a policy for modifying salaries based upon attendance, as well as firing teachers after the first year, in practice, no formal sanctions for less than contracted effort were applied during the adult education classes: no one was fired, pay was not reduced, no follow-up visits, etc. Sanctions (primarily firing decisions) were made six months after the end of classes.

⁸ Due to funding constraints, CRS could not introduce adult education into the control villages in 2016, but will plan to do so in 2017.

⁹ While we only have 20 villages in the control group, our power calculations were based upon previous research in Niger on adult education outcomes. Aker et al (2012) found that an adult education program that used mobile phones as a pedagogical tool in the classroom increased writing and math test scores by .20-.25 s.d. as compared with a traditional adult education program. The non-experimental before-after comparison of the traditional adult education program in that research formed the basis of the power calculations for this paper. We determined that a sample of 20 villages in the control group was sufficient to determine the causal impact of the adult education intervention.

¹⁰ During the baseline, several villages initially assigned to the program and research were removed, primarily due to internal village conflicts between two village chiefs. These were in the Zinder region. Thus, the final sample was 131 villages, with 111 villages in the adult education program, 57 in the mobile monitoring condition and 54 in the no mobile monitoring condition.

Individual-level eligibility was determined by two criteria: illiteracy (verified by an informal writing test) and willingness to participate in the adult education program. The same recruitment process during the baseline was followed for the control villages, as these villages were intended to receive the adult education program in 2016. Thus, our sample across all villages consists of those individuals who were selected by CRS to participate in the adult education program, regardless of whether or not they actually participated. Thus, the sampling composition is the same for both the adult education and pure control villages.

II. Model

A simple conceptual framework provides some intuition as to how monitoring might affect teachers' effort and student learning. A principal (the NGO or government) hires a short-term contractual teacher to teach an adult education program, but is unable to obtain complete information about the teachers' effort, related to imperfect supervision. Assuming that teachers believe they *may* be fired or penalized, monitoring should increase teachers' effort, which can vary with the intensity of monitoring and the cost of being fired.

Suppose that the NGO hires adult education teachers at a wage rate, w_{NGO} . If hired by the NGO, teachers can choose to exert some effort: $e=1$ (non-shirker) or $e=0$ (shirker). For simplicity, there are only two effort levels. Teachers who exert some effort will remain employed by the NGO for the duration of their contract. However, those who exert zero effort (shirkers) risk being caught (and fired) probability θ . These teachers can find a new job with probability p_m and receive an outside wage w_m , which requires effort e_m .

Using this framework, the utility function from shirking and non-shirking is therefore:

$$(1) \quad \begin{aligned} U^{NS} &= w_{NGO} - e \\ U^S &= (1 - \theta)w_{NGO} + \theta p_m (w_m - e_m) \end{aligned}$$

In order to extract positive levels of effort from the teachers, the NGO must choose a wage rate which assures that $U^{NS} \geq U^S$, or that the non-shirking condition is satisfied:¹¹

$$(2) \quad w_{NGO} \geq p_m(w_m - e_m) + \frac{e}{\theta}$$

The higher the teacher's outside option (outside wage net effort), the less likely he or she is to accept the NGO wage offer.¹² Assuming that the teacher accepts the NGO's offer, the teacher will then choose effort to maximize his/her expected utility.

Outside wage rates can vary by individual (w_m^i), as some teachers may be able to find a job more easily, based upon education, experience and gender. This will modify the non-shirker's utility function (slightly) to an individual-specific one, $U^{S,i}$. This suggests that the NGO should tailor the wage and monitoring to the teacher's outside options, but in practice, the NGO can only set a single wage, which will not satisfy the non-shirking condition for every teacher. As a result, a proportion of the teachers will shirk.

A mobile phone monitoring intervention affects the teacher's probability of being caught and fired θ , with $\theta_T \in (\theta_L, \theta_H)$, where L corresponds to the default (low monitoring) state and H to the additional mobile phone monitoring. Incorporating individual-specific outside options and treatment-specific firing probabilities into the utility function leads to the following modifications to the teacher's decision problem:

$$(3) \quad \begin{aligned} U^{NS} &= w_{NGO} - e \\ U^{S,i} &= (1 - \theta_T)w_{NGO} + \theta_T p_m(w_m^i - e_m) \end{aligned}$$

¹¹Whether or not teacher's effort (e) is influenced by the NGO wage rate (w_{NGO}), as in an efficiency wage model, would not affect the conclusions from our model. For simplicity, we abstract from this issue.

¹²In theory the NGO has two tools at its disposal to ensure teachers exert effort, namely w_{NGO} and θ , and the optimal combination of the two will be the outcome of the NGO's optimization process, including the cost of monitoring. Unless the wage is chosen such that no one shirks, the exact levels will not change our results.

Thus, the w_m^{i*} of the marginal teacher who is indifferent between working and shirking will depend upon the level of monitoring. Again, since the NGO cannot set an individual-specific wage rate, a proportion $\tau(w_{NGO}, \theta)$ of teachers will shirk.

Student learning outcomes are characterized by the following education production function:

$$(4) \quad y_i = y(e_i^t) \begin{cases} y(0) \text{ if } e = 0 \\ y(1) \text{ if } e = 1 \end{cases}$$

where e_i^t is the effort exerted by student i 's teacher, and teacher effort positively affects learning outcomes. This model does not show complementarities or substitutes between teacher and student effort. The average student outcome will therefore be a function of the share of teachers providing effort:

$$(5) \quad \bar{y} = \tau_T y(0) + (1 - \tau_T) y(1)$$

This leads to the following predictions with mobile phone monitoring:

- **Prediction 1.** As the probability of getting fired rises (θ_T), then $\frac{\partial U^S}{\partial \theta_T} < 0$, so $\frac{\partial \tau}{\partial \theta_T} > 0$. This is true whenever the NGO wage is greater than the outside wage net effort option, but this needs to be the case for teachers to accept the post in the first place. Since student achievement rises in teacher effort, then $\frac{\partial \bar{y}}{\partial \theta_T} > 0$
- **Prediction 2.** If the attractiveness of the teacher's outside option rises, i.e. p_m or $(w_m^i - e_m)$ rises, then the consequences of shirking become less severe and the proportion of teachers providing effort goes down: i.e. $\frac{\partial \tau}{\partial p_m} > 0$ and $\frac{\partial \tau}{\partial (w_m - e_m)} > 0$. This implies that students' learning outcomes will decrease with the attractiveness of teachers' outside options, so that $\frac{\partial \bar{y}}{\partial p_m} < 0$.¹³

¹³ This is not necessarily true when $p_m(w_m^i - e_m)$ and teacher ability are correlated, as then a higher-ability teacher might still teach better even when shirking than a "present" low-ability teacher. Then locally, the above result holds, but not when you change outside options in a discrete way. At this point the fact that we have measures of teacher ability become important. Conditional on ability the above results hold.

While this model focuses on the probability of being fired, in practice, the NGO did not use the mobile monitoring intervention to fire teachers during the adult education classes or between the first and the second year. Yet 23 percent of teachers were fired in a public and observable manner, with little information on the reasons. Thus, assuming that teachers believe they *may* be fired or penalized, additional monitoring should increase teachers' effort and student learning. Nevertheless, if teachers in mobile monitoring villages learn that there are no consequences to their shirking during the first year, the effects may dissipate during the second year.

IV. Data and Estimation Strategy

The data we use in this paper come from four primary sources. First, we conducted individual math and reading tests and use these scores to measure the impact of the program on educational outcomes. Second, we implemented household-level surveys. Third, we collected administrative and survey data on teachers, and use these data to better understand how the intervention affects teachers' effort. Fourth, we collected student attendance data from the centers in order to better understand if the intervention affected student attendance. Before presenting our estimation strategy, we discuss each of these data sources in detail.

A. Test Score Data

Our NGO partner identified students in all villages and for all cohorts in January 2014. While we had originally intended to implement the baseline in all 134 villages, the delayed start of the adult education program during the first year, as well as delays in funding, meant that we were only able to conduct the baseline in a subset of the sample (91 villages).¹⁴ In these villages, we stratified students by gender and took a random sample of 16 students per village, 11 women and 5 men. We implemented reading and math tests prior to the start of courses (February 2014), providing a baseline sample of approximately 1,271 students. We administered follow-up tests in the same baseline villages

¹⁴To choose the baseline villages, we stratified by region, sub-region and treatment status and selected a random sample of villages for the baseline.

(91) as well as the non-baseline villages in August 2014 and 2015, thereby allowing us to estimate the immediate impacts of the program. This total intended sample was 2,096 students, excluding attrition.

To test students' reading and math skills, we used USAID's Early Grade Reading Assessment (EGRA) and Early Grade Math Assessment (EGMA) tests. These are a series of individual tasks in reading and math, often used in primary school programs. EGRA is a series of timed tests that measure basic foundational skills for literacy acquisition: recognizing letters, reading simple words and phrases and reading comprehension (Dubeck and Gove 2015). Each task ranges from 60-180 seconds; if the person misses four answers in a row, the exercise is stopped. EGMA measures basic foundational skills for math acquisition: number recognition, comparing quantities, word problems, addition, subtraction, multiplication and division (Reubens 2009).

The EGRA and EGMA tests were our preferred survey instruments, as compared with the Ministry's standard battery of writing and math tests, for two reasons. First, most adult education programs are criticized for high rates of skills' depreciation. Yet these high rates of skills' depreciation may be simply due to the *automaticity* of reading achieved by the end of adult education programs, which are often not captured in traditional untimed tests. Since the short-term memory required to store deciphered material stores 7 items and lasts only 12 seconds, (Abadzi 2003). "Neoliterates must read a word in about 1-1.5 second (45-60 words per minute) in order to understand a sentence within 12 seconds (Abadzi 2003)."¹⁵ Thus, the EGRA timed tests allow us to determine whether participants in adult education classes are attaining the threshold required for sustained literacy acquisition. Second, the tests offer a great deal of precision in terms of measuring the skills that contribute to reading acquisition, such as simple decoding and reading comprehension (Dubeck and Gove 2015).

¹⁵This speed corresponds to oral-reading U.S. norms for first grade children. However, this is often not attained in literacy classes. For example, studies in Burkina Faso indicate that most literacy graduates need 2.2 seconds to read a word and are correct only 80-87 percent of the time (Abadzi 2003).

During the reading and math tests, we also measured students' self-esteem and self-efficacy, as measured by the Rosenberg Self-Esteem Scale (RSES) and the General Self-Efficacy Scale (GSES). The RSES is a series of statements designed to capture different aspects of self-esteem (Rosenberg 1965). Five of the statements are positively worded, while the other five statements are negatively worded. Each answer is assigned a point value, with higher scores reflecting higher self-esteem. The GSES is a ten-item psychometric scale that is designed to assess whether the respondent believes he or she is capable of performing new or difficult tasks and to deal with adversity in life (Schwarzer and Jerusalem 1995). The scale ranges in value from 12-60, with higher scores reflecting higher perceived self-efficacy. We use these results to measure the impact of the program on participants' perceptions of self-esteem and self-efficacy.

Survey attrition is a concern in most studies, especially in populations that engage in seasonal migration. Table A1 formally tests whether there is differential attrition by treatment status for the follow-up survey rounds. The rate of attrition in the comparison group was 5 percent in the first year, with relatively higher attrition in the normal adult education (without mobile monitoring) group and lower attrition in the mobile monitoring group. This suggests that the monitoring program might have prevented student attrition, at least in the first year. There was no differential attrition during the second year. Non-attriters in the adult education villages were more likely to be female as compared with non-attriters in the comparison villages, although there were no statistically significant differences among other characteristics. The difference in attrition by gender would likely bias our treatment effect for the adult education program downwards, as female students had lower test scores as compared with male students in adult education classes (Aker et al 2012). As a result, we bound our treatment effects using Lee bounds as a robustness check.

B. Household Survey Data

The second primary dataset includes information on household characteristics. We conducted a baseline household survey in February 2014 with 1,271 adult education students across 91 villages, the

same sample as those for the test score data, and a follow-up survey in all villages in December 2015. The survey collected detailed information on household demographics, assets, production and sales activities, access to price information, migration and mobile phone ownership and usage. These data are primarily used to test for baseline imbalances across the different treatments, as well as other impacts of the program.

C. Teacher Data

The third dataset is comprised of teacher-level characteristics and a measure of teachers' motivation. Using administrative data from CRS' teacher screening and training process, the dataset includes information on teachers' level of education, age, gender and village residence. In addition, we conducted a survey of all teachers in adult education villages, which included an intrinsic motivation inventory (IMI), in 2014 and 2015. The IMI is a multidimensional measurement instrument intended to assess participants' subjective experience related to a target activity, and has been used in several experiments related to intrinsic motivation and self-regulation (e.g., Ryan 1982, among others). The instrument assesses participants' interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension and perceived choice while performing a given activity, thus yielding six subscale scores that are combined into an overall score.¹⁶ We applied one of the versions of the IMI to our specific context, namely, teachers' experience in teaching the adult education program.

In addition to this teacher survey data, at the end of the adult education courses in May 2016, we also conducted classroom observations in five villages, using the modified Stallings classroom observation tool. While the sample size is too small to conduct statistical tests, we use these data to provide some supporting evidence as to whether the monitoring calls affected teaching quality.

¹⁶Although the overall questionnaire is called the IMI, the interest/enjoyment subscale is the only one that assesses intrinsic motivation. "The perceived choice and perceived competence concepts are theorized to be positive predictors of both self-report and behavioral measures of intrinsic motivation, and pressure/tension is theorized to be a negative predictor of intrinsic motivation." <http://selfdeterminationtheory.org/intrinsic-motivation-inventory/>.

Finally, we obtained data on teacher attendance collected by CRS in 2015, during the second year of the program. While these data were only for one year and for a subset of villages, they are used to provide a robustness check on teachers' self-reported attendance data in the survey.

D. Student Attendance

The final data are monthly student attendance data collected from a subset of intervention villages from CRS in 2015. These data are used to provide a “check” on teacher self-reported attendance, as well as to understand whether the interventions affected students' attendance within the classroom.

E. Pre-Program Balance

Table 1A shows the pre-program comparison of a number of student and household-level characteristics between the different treatments and control, controlling for the variables used for stratification (Bruhn and McKenzie 2009). Overall, the results suggest that the randomization was successful in creating comparable groups along observable dimensions. Differences in pre-program household characteristics are small and insignificant (Table 1, Panel A). Average age was 34, and a majority of respondents were members of the Hausa ethnic group. The average education level of household members was 2 years. Fifty-eight percent of households in the sample owned a mobile phone, with 61 percent of respondents having used a mobile phone in the months prior to the baseline. Respondents primarily used the mobile phone to make and receive calls. All respondents reporting *receiving* calls (as compared with making calls), as making a phone call requires being able to recognize numbers on the handset. While some baseline differences are statistically significant – such as asset and mobile phone ownership, which are related -- overall, we made over 100 baseline comparisons across the treatment groups.¹⁷ For each of the panels in Table 1A, we also test for the joint orthogonality of the covariates, with p-values of .25, .48 .55 and .11, respectively.

¹⁷The dependent variable in these regressions is “monitor” and is only estimated on the subset of adult education villages, so tests for the joint orthogonality of covariates with respect to assignment to the monitoring treatment.

Table 1B provides further evidence of the comparability across treatments for reading scores. Using non-normalized baseline reading scores for each task, students in control villages (i.e., without the adult education program) had low levels of letter, syllable and word recognition prior to the program, with students in control villages being able to correctly identify 2 letters, 1 syllable and 1 word. There was not a statistically significant difference in these baseline reading levels between the adult education and control groups or between the mobile monitoring and non-mobile monitoring villages. Baseline math scores suggest that students had relatively higher levels of math knowledge, as they were able to correctly identify 4-5 numbers prior to the program (Table 1C). These baseline math levels were similar across the comparison, adult education and mobile monitoring groups. Overall, the baseline comparisons suggest that the project successfully selected participants who were functionally non-literate prior to the start of the program, and that baseline levels were similar across groups.

Table 1D presents a comparison of teacher characteristics across the adult education villages. Overall teacher characteristics are well-balanced between the mobile monitoring and non-mobile monitoring villages. Teachers in adult education classes without mobile monitoring were 37 years old and 35 percent had some secondary education. Roughly one-third of the adult education teachers were female, and a strong majority were married.

D. Estimation Strategy

To estimate the impact of both the adult education program and mobile monitoring on educational outcomes, we use a simple differences specification. Let $test_{iv}$ be the reading or math test score attained by student i in village v immediately after the program in 2014 and 2015.¹⁸ $adulterd_v$ is an indicator variable for whether the village v is assigned to the adult education intervention ($adulterd=1$) or the control ($adulterd=0$). $adulterd*monitor_i$ takes on the value of one if the adult

¹⁸There are a number of ways that raw EGRA and EGMA scores can be used and transformed for analysis, including the raw untimed scores, raw timed scores (especially for reading scores), untimed normalized scores and timed normalized scores. The results in the tables show the timed normalized scores for reading and the untimed normalized scores for math, as is the convention for EGRA and EGMA. Results are largely robust to using raw non-normalized scores.

education village received the mobile monitoring intervention, and 0 otherwise. θ_S are geographic fixed effects at the regional and sub-regional levels (the level of stratification). We pool observations across the two years and estimate the following specification:

$$(6) \quad test_{iv} = \beta_0 + \beta_1 adulated_v + \beta_2 adulated_v * monitor_v + \theta_S + \varepsilon_{iv}$$

The coefficients of interest are β_1 and β_2 , which capture the average immediate impact of the adult education program (without monitoring) and the additional impact of the mobile phone monitoring intervention. The error term ε_{iv} captures unobserved student ability or idiosyncratic shocks. We cluster the error term at the village level for all specifications and add in a binary variable for the second year as a robustness check.¹⁹

Equation (6) is our preferred specification. As a robustness check to this preferred approach, we also estimate the impact of the program using a value-added specification and using alternative measures of the dependent variable. However, the value-added specification reduces our sample size considerably, as we only have baseline data for 91 villages.

V. Results

Figures 2A and 2B depict the mean normalized reading and math test scores by treatment status. Test scores are normalized using the mean and s.d. of contemporaneous test scores in control villages.²⁰ Three things are worth noting. First, the adult education program increases reading and math scores significantly as compared to the control group, with relatively stronger effects on reading. Second, for reading, the impacts of the adult education program are stronger for simpler “decoding” tasks, namely, letter or syllable recognition. For math, however, the impacts of the adult education program are stronger for more difficult math tasks, such as addition, subtraction, multiplication and division, as some students were able to recognize numbers prior to the program.²¹ And third, the

¹⁹All results are robust to including a binary variable for the second year. Results are available upon request.

²⁰Normalizing the z-scores by region also yields similar results.

²¹Students in adult education villages without mobile monitoring did worse in quantity comparison as compared with the control group. This task asks students to compare a set of two numbers with three digits that closely resemble each other.

difference in test scores between mobile monitoring and non-mobile monitoring villages represents 60% of the difference in test scores between the standard adult education villages and the control group, especially for simpler reading and math tasks. This suggests important learning gains from the mobile monitoring program.

A. Pooled Impacts of the Intervention

Table 2 presents the results of Equation (3) for reading z-scores across both years of the program. Across all reading tasks, the adult education intervention alone increased students' reading test scores by .14-.30 s.d. over the two-year period, with statistically significant effects (Table 2, Panel A). Similar to Figure 2, the impacts are stronger for simpler decoding tasks and the composite reading score. By the end of the program, students in the standard adult education classes could read approximately four more letters, two more syllables and two more words as compared to those in the control villages (Table A2). Nevertheless, the program did not raise students' reading scores to threshold reading level of 1 word per 1.5 seconds.

The mobile monitoring intervention increased reading z-scores by an additional .07-.19 s.d., with a statistically significant effect at the 5 and 10 percent levels for four of the six reading measures. Similar to the adult education results, the monitoring impacts were stronger in magnitude and statistical significance for simpler reading tasks – i.e., decoding letters, syllables and words – as compared to phrases and reading comprehension. By the end of the program, students in the mobile monitoring villages were able to read an additional two letters and syllables and an additional word as compared to those in the standard adult education villages (Table A2).

Psychology shows that individuals use simple cognitive shortcuts when processing information, such as “left digit bias”, or the tendency to focus on the left-most digit of a number while partially ignoring other digits (Lacetera et al 2012). Whereas those who are non-literate guess for this task (and are correct 50% of the time), those who are neo-literate try. However, if they are not yet literate enough to process all of the information provided, we posit that they left-most digit (ignoring other numbers) and thus may guess incorrectly. This task is designed to text for this, as most of the number start with the same left-hand digit.

The results are similar for math z-scores (Table 3): the adult education program increased math z-scores by .13-.29 s.d. as compared with the control group (Panel B, Column 1), with statistically significant effects at the 1 and 5 percent levels for more difficult math tasks and the composite score.²² With the exception of simple quantity comparison, the program was successful in moving students beyond simple number identification to completing more complicated mathematical tasks, namely addition, subtraction, multiplication and division. Concretely, this means that students were able to correctly complete an additional addition and subtraction problem by the end of the program.

The mobile monitoring intervention further increased math z-scores by .07-.12 s.d., with statistically significant effects primarily for number identification and quantity comparison (Panel A, Columns 1 and 2). While the adult education program was successful in increasing math z-scores for most tasks, with the exception of quantity comparisons, the monitoring impacts were primarily stronger for number identification and quantity comparisons.

B. Effects of the Program over Time

While the results in Tables 2 and 3 suggest that the mobile monitoring intervention increased reading and math z-scores as compared with the standard adult education program, a key question is the dynamics of these effects over time, once teachers learned more about the monitoring intervention and adults achieved higher learning outcomes.

Tables 4 and 5 shows the results of the adult education and mobile monitoring interventions by year. Two things are worth noting. First, math and reading z-scores increase in the adult education villages over time, but are stronger for math.²³ Second, the effects of the mobile monitoring intervention are positive and statistically significant in the first year, primarily for the simpler reading and math tasks and the composite scores (Tables 4 and 5, Panel A). Yet the coefficients for the

²²While math tasks in the EGMA tests are also timed, most analyses use untimed scores in the analysis. An interesting impact is the negative coefficient on the “quantity comparison” task; this task essentially involves asking students to compare similar three-digit numbers and note which one is larger, such as 997 and 979.

²³This is also the case if we pool the 2014 and 2015 data and include a time trend interacted with both the adult education and monitoring treatments. While the coefficient on the interaction terms for the adult education and time trend are positive, they are primarily statistically significant for math.

monitoring intervention are not statistically significant for any of the reading tasks in the second year, and for only one of the math tasks.²⁴

Does this lack of statistical significance in the second year suggest that the intervention was less effective due to teachers' learning about the intervention? While the coefficients on the monitoring intervention over time are not statistically significant, they are large in magnitude, representing 30-70% of the coefficients on the adult education program. Given the higher test scores during the second year, the learning effects due to monitoring would have had to increase significantly during the second year in order to detect an effect, which may be difficult if there are non-linearities in learning outcomes. In addition, as mentioned above, the monitoring intervention was modified slightly in the second year, so that a subset of villages received the full monitoring intervention.

We partially address this issue by restricting the sample only to villages that received the full monitoring intervention in 2015, similar to 2014 (Tables 4 and 5, Panel C). Using the restricted sample, and unsurprisingly, the coefficients on the impact of adult education program in 2015 are the same as in the full sample. For reading, the coefficients on the monitoring intervention are greater than or equal to the monitoring coefficients in the full sample, although still not statistically significant (Tables 4 and 5, Panel C). For math, the full monitoring intervention increased math z-scores in 2015, with a statistically significant effect for quantity comparisons and multiplication and division z-scores. These were two tasks for which there were no monitoring effects in the first year, and which are traditionally more difficult for beginning students.

Taken together, these results suggest that the monitoring intervention had an impact on learning in the second year, although primarily for math tasks, suggesting that teachers in the monitoring villages did not completely reduce their effort in the second year. Nevertheless, the results also

²⁴Estimating the regressions jointly with an interaction between the adult education, monitoring and time fixed effects supports these results.

suggest that full monitoring intervention was necessary to improve learning, as opposed to only calling the teachers.²⁵

C. Heterogeneous Effects by Region, Gender and Teachers' Characteristics

We might expect greater impacts of the monitoring intervention among certain sub-populations or according to teachers' characteristics, as predicted by our model. Table 6 tests for heterogeneous impacts of the program by the student's geographic location and gender, while Table 7 tests for heterogeneous effects by teacher characteristics.

Student Characteristics

While the Zinder and Maradi regions both primarily engage in agriculture and livestock, the Zinder region is relatively closer to Nigeria, with relatively higher baseline reading and math test scores. At the same time, the Zinder region had a greater number of villages. Although the number of villages per field agent were similar across regions, the monitoring program could be more useful in the Zinder region if it was more difficult to travel to villages. Columns 1 and 2 report the results of a triple difference-in-differences (DDD) regression that tests for differential effects of the monitoring program by region. The triple interaction term is not statistically significant for reading or math z-scores, suggesting that the monitoring program did not have a differential impact by region.

In light of different socio-cultural norms governing women's and men's household responsibilities and social interactions, the adult education and monitoring program could have differential impacts by students' gender. As women belonging to particular ethnic groups in Niger travel outside of their home village less frequently than men, women may have had fewer opportunities to practice their newly-acquired skills outside of class. In addition, given the larger student-to-teacher

²⁵An alternative mechanism to test whether the way to test whether the smaller effect in the second year is due to teacher learning is to identify teachers who shirked during the first year, but who were not fired. In theory, these teachers should exert less effort in the second year, as they were not sanctioned, and therefore have lower test scores. Nevertheless, only 25% of teachers were fired between the first and second year, and we do not have similar shirking information on non-monitoring teachers. Restricting the sample to retained teachers and the full monitoring intervention in the second year, however, we find similar results: positive and not statistically significant results for reading, and statistically significant results for some math tasks.

ratio in women's classes in the program, this could have negatively affected women's learning outcomes. Overall, women in the control group had significantly lower reading and math scores than men in control villages, confirming the "gender gap" in education skills in Niger. The adult education program increased men's reading and math z-scores by .11-.56 s.d., with relatively stronger effects on for simpler reading tasks and math skills. While women's reading and math z-scores were lower than men's, the results are not statistically significant. The monitoring component had a positive impact on men's test scores, primarily for reading, although these impacts are not statistically significant for most tasks. Overall, the triple interaction shows that the monitoring effects do not differ by gender. This is perhaps unsurprising, as the same teacher taught both men's and women's classes in the village. At the same time, since women's class sizes were larger than men's, we are unable to disentangle the "gender" effect from the "class size" effect.

Teacher Characteristics

Table 7 presents the impact of the mobile monitoring program on reading and math composite z-scores by teachers' characteristics, such gender, education, previous experience as an adult education teacher and whether the teacher lives within the village.²⁶ In many villages in the Maradi and Zinder regions, women rarely migrate outside of the village for work; as a result, female adult education teachers might have fewer outside options, thereby making the monitoring component more effective. This is confirmed by the teacher survey: While 46 percent of male teachers reported that they could find other work if they were not adult education teachers, only 24 percent of female teachers did so, despite the fact that both groups had similar education levels. On the other hand, teachers with higher levels of education should have better outside options, thereby reducing the effectiveness of monitoring component.

Overall, the monitoring program was associated with positive improvements in reading and math z-scores for male teachers, increasing reading and math z-scores by .11-.15 s.d. as compared with

²⁶All regressions are conditional on the presence of an adult education program in the village.

the non-mobile monitoring villages (Columns 1 and 5). The effect was stronger for female teachers – an additional .12-.17 s.d. -- although not statistically significant at conventional levels. This suggests that the monitoring intervention was slightly more effective for teachers for whom the outside option was relatively lower, although does not provide conclusive evidence.

While teachers' education levels did not have a strong effect on learning levels – either in standard or mobile monitoring villages - teachers' experience is negatively correlated with learning outcomes (Columns 3 and 7), suggesting that these teachers are not putting in the same level of effort as newer teachers.²⁷ Yet the monitoring intervention somewhat mitigates this effect. While initially surprising, seasoned adult education teachers have experience in a “niche” market, and have been outside of the traditional migration labor force – which coincides with the period of the adult education classes – - for several years. As such teachers have lower outside options, as is supported by their self-reports during the teacher survey, the monitoring intervention had a stronger effect.

Finally, a key question is whether the monitoring intervention has a stronger effect on “local” or more distance teachers. While the intervention could potentially make it easier for the community to observe teachers' absence – especially for those who are traveling from outside villages – the nature of the intervention may be more effective for local teachers, as they are subject to immediate social pressures within the community. Columns (4) and (8) suggest that the latter scenario is the case: While local teachers had higher reading and math z-scores than their non-local counterparts, the monitoring intervention had a strong effect on local teachers, increasing students' test scores by .23-.34 s.d. This suggests that the monitoring intervention allowed the community to put greater pressure on teachers, but primarily when they were closer to the community.

VI. Potential Mechanisms

²⁷Average teacher experience is 2.5 years, with a s.d. of 5 years.

There are a variety of mechanisms through which the mobile monitoring intervention could affect students' learning. First, mobile monitoring can potentially lead to increased teacher effort, reflected by reduced absence or improved classroom pedagogy, thereby improving the effectiveness of the overall adult education curriculum. Second, the phone calls could potentially increase teachers' intrinsic motivation, thereby increasing their teaching efficacy within the classroom. Third, having a more present and motivated teacher could potentially affect students' effort, leading to increased class participation and attendance. Fourth, as the monitoring component involved a subset of students in the class, the calls could have motivated students independently, thereby leading to spillover effects on their fellow learners. And finally, since the monitoring component also involved village chiefs, this could have increased their interest in community-level development programs, thereby motivating teachers and students. We present evidence on each of these mechanisms in turn.

A. Teacher Effort and Motivation

The mobile phone monitoring could have increased teacher effort within the classroom and students' performance in two ways. First, it could have encouraged teachers to teach more classes. Second, it could have improved teachers' efficacy within the classroom, thereby improving student learning. In fact, the teacher surveys report to both of these effects: Teachers reported that "The...calls prevent us from missing courses", and that "Someone who works must be 'controlled'", as well as that "The (calls) prove that our work is important."

Yet in the absence of financial punishments or rewards, in order for these mechanisms to hold, teachers needed to believe that the monitoring calls were a review of their performance and that some reward or punishment was possible. Overall, 70% of teachers thought the calls were from CRS, whereas 29% from the Ministry, suggesting that they understood that the calls were from supervisory figures. In addition, within the full monitoring villages, teachers were generally aware of the other monitoring calls: 80% of teachers reported knowing that the village chief was called, and 77% reported knowing that some students were called. The monitoring intervention did not appear to have

a strong impact on teachers' presence within the classroom, although these indicators were imperfectly measured.

In order to test the above mechanisms, we would ideally have high-frequency data on teacher absence from all villages. However, due to limited in-person monitoring by CRS and the nature of the mobile monitoring intervention, we only have such information from mobile monitoring villages.²⁸ As a result, we assess the impact of the monitoring intervention on teacher effort using a number of proxies. To measure teacher absence, we use self-reported measures, as well as attendance measures collected by the Ministry. Table 8 shows the results of the monitoring component on these indicators. While 53 percent of non-monitoring teachers reported stopping the course at some time – on average for 2 days -- teachers in monitoring villages were only slightly less likely to do so, and reported being absent for 1.30 fewer days than the non-monitoring teachers (Panel A). The primary reasons cited for absence in both monitoring and non-monitoring villages were illness, funerals and agricultural work. While this suggests that the monitoring intervention may have reduced teacher absenteeism, these indicators are self-reported and quite small in magnitude; as a result, they could not solely explain the improvements in learning.²⁹ This is confirmed, in part, by CRS' attendance records: non-monitoring teachers taught an average of 19.25 days a month, and monitoring teachers were not more likely to teach more classes (Panel B). This is supported by the teacher firing data: There was no correlation between mobile monitoring and the teacher's likelihood of being replaced between the first and second year. While these null results could, in part, be explained by the noisiness of the attendance data and

²⁸A potential critique of the mobile monitoring intervention is that the observed changes are simply due to the Hawthorne effect; in other words, teachers are changing their behavior (and increasing their effort) simply because they are being monitored. This is the precise purpose of the intervention and, we would argue, something that is inherent in all monitoring interventions: The purpose of such interventions is to reduce the information asymmetry between the principal and agent.

²⁹Despite the fact that these data were self-reported, there was a high intra-village correlation of responses amongst teachers, village chiefs and students in monitoring villages, even when the teacher was absent, and this did not appear to change over time (potentially because different students were called). While this could be due to either collusion or a high degree of information-sharing amongst the stakeholders, CRS did not use the monitoring data to make firing decisions between the first and the second year.

the relatively smaller sample size of those data, the monitoring intervention did not appear to have strong effects on teacher absenteeism.³⁰

The calls could have affected teachers' motivation, thereby making them more effective in class, if the intervention did not crowd out intrinsic motivation.³¹ We proxy motivation in two primary ways: an observable measure of teacher "additional" effort and the intrinsic motivation inventory (IMI). For the former measure, CRS and the Ministry suggested that teachers keep attendance logs as a means of better managing their classroom, but did not require such logs or verify them. While 28 percent of non-monitoring teachers kept their own attendance logs, monitoring teachers were 18 percentage points more likely to do so, with a statistically significant effect at the 5 percent level. This suggests that monitoring teachers were more willing to invest in teaching preparation (Panel C). To measure intrinsic motivation, we use several sub-scales of the IMI, namely intrinsic motivation, perceived competence, perceived pressure and perceived choice. While the monitoring intervention did not have an impact on teachers' perceived competence, pressure or perceived choice, monitoring teachers had intrinsic motivation z-scores that were .24 s.d. higher than their non-monitoring counterparts, with a statistically significant effect at the 10 percent level. While self-reported, this suggests that the monitoring intervention increased teachers' motivation vis-à-vis the teaching tasks.

B. Student Effort and Motivation

The monitoring component could have encouraged greater student effort within the classes, as measured by student dropout, attendance and motivation. While we do not have reliable data on student attendance, we do have self-reported measures of student dropout, the reasons for dropping out and the duration of time in the course. Table 9 shows these results. Overall, 27% of students dropped out of the course at some point in the time over the two-year period, and the monitoring component did not affect the likelihood of student dropout (Panel A). Since a majority of those who dropped out

³⁰Of the adult education villages, CRS only had attendance records for a subset of villages.

³¹We attempted to collect classroom observation data using a modified Stallings observation tool in May 2016, after the end of the classes, we were only able to go to a small number of villages (5) before adult education classes ended for the season.

primarily did so for reasons outside of their control, namely, pregnancy, illness or a death in the family, the lack of an observed impact is perhaps not surprising. For those who stayed in the course, the monitoring intervention did not appear to affect how long students stayed in the course. This suggests that students were not necessarily spending more time in the course.

There is some evidence that the monitoring component affected student learning via the mechanism of calling students directly. Panel B shows the results of a regression of test scores on a binary variable for students who were called, as well as the monitoring treatment and an interaction term between the two.³² The intervention appeared to affect the “called” students’ learning outcomes: called students had significantly higher reading z-scores as compared with non-called students in monitoring villages, as well as students in non-monitoring villages. It is possible that the called students’ greater motivation passed to other students, although we cannot directly test this hypothesis.³³

VII. Alternative Explanations

There are several potential confounds to interpreting the above findings. First, there might be differential in-person monitoring between monitoring and non-monitoring villages. If the Ministry of Non-Formal Education or CRS decided to focus more of their efforts on monitoring villages because they had better information, then any differences we observe in test scores might be due to differences in program implementation, rather than the monitoring component. Yet during the implementation of program, there was very little in-person monitoring, and no differential visits by treatment status.

A second potential confounding factor could be due to differential attrition. The results in Table A1 suggest that attrition is higher in the adult education villages as compared with the comparison group and lower in the monitoring villages (as compared with non-monitoring villages), primarily during the first year. Women are slightly more likely to remain within the sample in adult education

³² The results in Table 9 (Panel B) excludes the control villages.

³³ The main results are robust to excluding the “called” students from the sample, although the magnitudes of the coefficients are smaller (Table A3).

villages (without monitoring) as compared to the control. Since women had lower reading and math z-scores overall, this may underestimate the effects of the adult education program alone. By contrast, there are fewer women within the monitoring villages (as compared with the non-monitoring villages), which could potentially overestimate the effects of the monitoring program as compared with the adult education program. As we are primarily concerned with this latter comparison, we use Lee bounds to correct for bias for differential attrition between the monitoring and non-monitoring villages across both years. Table A4 shows that the upper bounds remain positive and statistically significant for all task, and that the lower bounds for reading and math z-scores are generally positive, but not statistically significant for most tasks.³⁴

Finally, for some of the student and teacher survey measures, there could be concerns about non-classical measurement error, as teachers and students could systematically report in ways that would bias the results. While this is an obvious concern for self-reported attendance data, when possible, we attempted to verify these results with administrative data. For the student test score data, as these are short, timed tests that objectively measure students' learning, and cannot be easily manipulated, we are less concerned about this potential issue.

A final potential confounding factor could emerge if the monitoring intervention is not pure monitoring intervention, but rather a “reminder” intervention. This could encourage students or teachers to prepare more for classes, thereby improving test scores. While this effect would still be attributed to the mobile monitoring program, it would have different implications for replicating the program: one interpretation would suggest a “monitoring” effect, whereas the other would suggest a “reminder” effect. Unfortunately, we cannot test for this empirically.

VIII. Cost-Effectiveness

³⁴The small number of observations in the comparison group who did not receive the adult education intervention could raise concerns that our confidence intervals are too narrow (Cameron, Gelbach and Miller 2008). However, when estimating the relative impact of the mobile monitoring program as compared with the adult education program (conditional on being in the adult education program), the results are similar.

A key question is the cost-effectiveness of the mobile intervention as compared to regular monitoring. While in-person monitoring visits were limited over the duration of the study, we have data on per-monitoring costs for both in-person and mobile monitoring (Figure 3). On average, in-person monitoring costs are \$6.20 per village, primarily including costs for the agent's time and gas for the motorcycle. By comparison, the mobile monitoring intervention only costs \$3.08 per village, including the costs of agents' time and mobile phone credit. This suggests that per-village savings are \$3, as compared with average gains of .07-.18 s.d. in learning over the two-year period

IX. Conclusion

Adult education programs are an important part of the educational system in many developing countries. Yet the successes of these initiatives have been mixed, partly due to the appropriateness of the curriculum, the opportunity costs of adults' time and the ability of governments and international organizations to monitor teachers' effort, who are often located in remote rural areas

This paper assesses the impact of an intervention that conducted mobile monitoring of as part of an adult education intervention in Niger. We find that simply monitoring teachers substantially increased students' skills acquisition over the two-year period, suggesting that mobile telephones could be a simple and low-cost way to improve adult educational outcomes. The treatment effects are striking although dynamic over time: the adult education program with monitoring increased reading and math test scores by .07-.18 s.d. as compared with the standard adult education program, with relatively stronger effects for reading in the first year and math in the second year. The impacts appear to operate through increasing teacher and student motivation within the classroom, and are primarily derived from the monitoring model that uses community pressure – i.e., by calling the village chief, the teacher and two students – rather than only the teacher.

References

- Abadzi, Helen.** 1994. "What We Know About Acquisition of Adult Literacy: Is There Hope?," In *World Bank discussion papers*, ix, 93 p. Washington, D.C.: The World Bank.
- Abadzi, Helen.** 2013. Literacy for All in 100 Days? A research-based strategy for fast progress in low-income countries," *GPE Working Paper Series on Learning No. 7*
- Aker, Jenny C., Christopher Ksoll and Travis J. Lybbert.** October 2012. "Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger." *American Economic Journal: Applied Economics*. Vol 4(4): 94-120.
- Aker, Jenny C., Christopher Ksoll, Danielle Miller, Karla Perez, Susan L. Smalley.** 2015. "Learning without Teachers? Evidence from a Randomized Experiment of a Mobile Phone-Based Adult Education Program in Los Angeles." *CGD Working Paper 368*.
- Andrabi, Tahir, Jishnu Das, Asim Ijaz Khwaja, and Tristan Zajonc.** 2011. "Do Value-Added Estimates Add Value? Accounting for Learning Dynamics." *American Economic Journal: Applied Economics*, 3(3): 29–54.
- Angrist, Joshua D., and Jorn - Steffen Pischke.** 2009. *Mostly Harmless Econometrics*:
- Banerjee, Abhijit, and Esther Duflo.** 2006. "Addressing Absence." *Journal of Economic Perspectives* 20 (1): 117–32.
- Banerjee, Abhijit, Shawn Cole, Esther Duflo and Leigh Linden.** 2007. "Remedying Education: Evidence from Two Randomized Experiments in India." *The Quarterly Journal of Economics*, 122(3), pp. 1235-64.
- Banerji, Rukmini, James Berry and Marc Shotland.** 2013. "The Impact of Mother Literacy and Participation Programs on Child Learning: A Randomized Evaluation in India."
- Barrow, Lisa, Lisa Markman and Cecilia Elena Rouse.** 2009. "Technology's Edge: The Educational Benefits of Computer-Aided Instruction." *American Economic Journal: Economic Policy*, 1(1), pp. 52-74.
- Blunch, Niels-Hugo and Claus C. Pörtner.** 2011. "Literacy, Skills and Welfare: Effects of Participation in Adult Literacy Programs." *Economic Development and Cultural Change*. Vol. 60, No. 1 (October 2011): 17-66.
- Bruhn, Miriam, and David McKenzie.** 2009. "In Pursuit of Balance: Randomization in Practice in Development Field Experiments." *American Economic Journal: Applied Economics*, 1(4): 200-232
- Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller.** 2008. "Bootstrap-based improvements for inference with clustered errors." *Review of Economics and Statistics* 90.3: 414-427.

- Carron, G.** 1990. "The Functioning and Effects of the Kenya Literacy Program." *African Studies Review*, pp. 97-120.
- Cilliers, Jacobus, Ibrahim Kasirye, Clare Leaver, Pieter Serneels, and Andrew Zeitlin.** 2014. "Pay for locally monitored performance? A welfare analysis for teacher attendance in Ugandan primary schools."
- Cueto, Santiago, Máximo Torero, Juan León, and José Deustua.** 2008. Asistencia docente y rendimiento escolar: el caso del programa META "Teacher support and school accountability: The META program". GRADE Working Paper No. 53. Lima, Peru: GRADE.
- De Ree, Joppe, Karthik Muralidharan, Menno Pradhan and Halsey Rogers.** 2016. "Double for Nothing? Experimental Evidence on the Impact of an Unconditional Salary Increase on Student Performance in Indonesia."
- DiNardo, J., J. McCrary, and L. Sanbonmatsu.** 2006. "Constructive Proposals for Dealing with Attrition: An Empirical Example." Working paper, University of Michigan.
- Dubeck, Margaret M. and Amber Gove.** 2015. "The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations." *International Journal of Educational Development* 40 (2015) 315–322.
- Duflo, Esther, Rema Hanna and Stephen Ryan.** 2012. "Incentives Work: Getting Teachers to Come to School," *American Economic Review*.
- Duflo, Esther.** 2012. "Women Empowerment and Economic Development." *Journal of Economic Literature*. 50(4). 1051-1079.
- Duflo, Esther, Pascaline Dupas and Michael Kremer.** 2015. "School Governance, Pupil-Teacher-Ratios, and Teacher Incentives: Experimental Evidence from Kenyan Primary Schools". *Journal of Public Economics* Vol. 123, pp. 92-110.
- Doepke, Mathias and Michele Tertilt.** 2014. "Does Female Empowerment Promote Economic Development?" *NBER Working Paper 19888*, NBER, Inc.
- Glewwe, Paul, and Michael Kremer.** 2006. "Schools, Teachers, and Education Outcomes in Developing Countries." In *Handbook of the Economics of Education* Volume 2, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 945–1017. Amsterdam: Elsevier.
- Guerrero, Gabriela, Juan Leon, Mayli Zapata & Santiago Cueto.** 2013. "Getting teachers back to the classroom. A systematic review on what works to improve teacher attendance in developing countries." *Journal of Development Effectiveness*. 5(4).
- Lee, David S.** 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects" *The Review of Economic Studies*, 6, 1072-1102.
- Muralidharan, Karthik, Jishnu Das, Alaka Holla and Aakash Mohpal.** 2014. "The Fiscal Cost of Weak Governance: Evidence from Teacher Absence in India." Unpublished mimeo.

- Ortega, Daniel and Francisco Rodríguez.** 2008. "Freed from Illiteracy? A Closer Look at Venezuela's Mision Robinson Literacy Campaign." *Economic Development and Cultural Change*, 57, pp. 1-30.
- Osorio, Felipe, and Leigh L. Linden.** 2009. "The use and misuse of computers in education: evidence from a randomized experiment in Colombia." *The World Bank Policy Research Working Paper Series*.
- Oxenham, John, Abdoul Hamid Diallo, Anne Ruhweza Katahoire, Anne Petkova-Mwangi and Oumar Sall.** 2002. *Skills and Literacy Training for Better Livelihoods: A Review of Approaches and Experiences*. Washington D.C.: World Bank.
- Reubens, Andrea.** 2009. "Early Grade Mathematics Assessment (EGMA): A Conceptual Framework Based on Mathematics Skills Development in Children." USAID: Washington, D.C.
- Romain, R. and L. Armstrong.** 1987. *Review of World Bank Operations in Nonformal Education and Training*. World Bank, Education and Training Dept., Policy Division.
- Ryan, R. M. (1982).** Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, 43, 450-461.
- Transparency International.** 2013. *The Global Corruption Report: Education*. Routledge: New York, NY.
- UNESCO.** 2005. *Education for All: Global Monitoring Report. Literacy for Life*. Paris: UNESCO.
- UNESCO.** 2008. *International Literacy Statistics: A Review of Concepts, Methodology and Current Data*. Montreal: UNESCO Institute for Statistics.
- UNESCO.** 2012. *Education for All: Global Monitoring Report. Youth and Skills: Putting Education to Work*. Paris: UNESCO.

Table 1A. Baseline Household Characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
	Comparison Group	Monitoring	Adult Educ.	Difference	Difference	p-value
	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Coeff (s.e)	Coeff (s.e.)	
<i>Household Characteristics at Baseline</i>				(2)-(1)	(3)-(1)	(2)=(3)
Age of Respondent	35.6 (12.98)	33.44 (11.63)	34.08 (12.01)	-1.26 (1.083)	-1.97 (1.273)	0.73
Gender of Respondent (1=Female, 0=Male)	0.685 (0.466)	0.677 (0.468)	0.683 (0.465)	0.01 (0.0121)	-0.01 (0.0217)	0.40
Average education level of household (in years)	1.787 (0.963)	2.112 (1.028)	2.069 (0.985)	0.12 (0.0811)	-0.08 (0.0906)	0.19
Number of asset categories owned by household	5.585 (1.543)	5.895 (1.6)	5.81 (1.569)	0.22* (0.115)	-0.15 (0.206)	0.16
Household experienced drought in past year (0/1)	0.471 (0.501)	0.564 (0.496)	0.537 (0.499)	0.03 (0.0400)	0.02 (0.0611)	0.83
Household owns a mobile phone (0/1)	0.58 (0.496)	0.685 (0.465)	0.665 (0.472)	0.07** (0.0339)	0.00 (0.0519)	0.33
Respondent used a cell phone since the last harvest	0.61 (0.502)	0.647 (0.478)	0.644 (0.479)	0.03 (0.0330)	0.03 (0.0577)	0.95
Used cellphone in past two weeks to make calls	0.737 (0.446)	0.722 (0.449)	0.703 (0.457)	0.04 (0.0338)	-0.05 (0.0591)	0.25
Used cellphone in past two weeks to receive calls	1 (0)	0.967 (0.178)	0.965 (0.185)	0.00 (0.0165)	-0.05*** (0.0227)	0.19

Note: This table shows the difference in means between the different treatment groups. "Comparison" is defined as villages assigned to no adult education treatment in 2014 or 2015. "Adult education" is defined as those villages that were assigned to adult education without monitoring, whereas "Monitoring" is defined as villages that were assigned to adult education with monitoring. Standard deviations are shown in parentheses. Columns (4) and (5) show the coefficients and s.e. from a regression of each characteristic on the treatments and stratification fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 1B. Baseline Reading Test Scores						
	(1)	(2)	(3)	(4)	(5)	(6)
	Comparison Group	Monitoring	Any Adult Educ.	Difference	Difference	p-value
	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	Coeff (s.e)	Coeff (s.e.)	
				(2)-(1)	(3)-(1)	(2)=(3)
Letter recognition	2.074 (7.115)	3.368 (10.71)	3.146 (10.29)	0.237 (0.667)	0.383 (0.632)	0.895
Syllable recognition	1.2 (5.532)	2.745 (9.754)	2.483 (9.362)	0.387 (0.611)	0.712 (0.480)	0.727
Word recognition	0.968 (5.17)	1.664 (7.277)	1.547 (7.299)	0.0762 (0.446)	0.155 (0.427)	0.914

Note: This table shows the difference in means between the different treatment groups. "Comparison" is defined as villages assigned to no adult education treatment in 2014 or 2015. "Adult education" is defined as those villages that were assigned to adult education without monitoring, whereas "Monitoring" is defined as villages that were assigned to adult education with monitoring. Standard deviations are shown in parentheses. Columns (4) and (5) show the coefficients and s.e. from a regression of each characteristic on the treatments and stratification fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 1.C. Baseline Math Test Scores						
	(1) Comparison Group	(2) Monitoring	(3) Any Adult Educ.	(4) Difference Coeff	(5) Difference Coeff	(6) p- value
	Mean (s.d.)	Mean (s.d.)	Mean (s.d.)	(s.e) (2)-(1)	(s.e.) (3)-(1)	(2)=(3)
Highest number correctly counted to	44.07 (23.75)	41.89 (24.24)	41.67 (23.95)	1.218 (1.576)	-0.963 (4.832)	0.677
Numbers correctly identified (out of 12)	4.135 (5.32)	4.414 (5.268)	4.342 (5.202)	0.122 (0.294)	0.217 (0.645)	0.899
Numbers correctly identified (out of 20)	5.708 (8.168)	5.791 (8.137)	5.747 (8.094)	-0.0105 (0.495)	0.105 (0.691)	0.906

Note: This table shows the difference in means between the different treatment groups. "Comparison" is defined as villages assigned to no adult education treatment in 2014 or 2015. "Adult education" is defined as those villages that were assigned to adult education without monitoring, whereas "Monitoring" is defined as villages that were assigned to adult education with monitoring. Standard deviations are shown in parentheses. Columns (4) and (5) show the coefficients and s.e. from a regression of each characteristic on the treatments and stratification fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 1D. Balance Table of Teacher Characteristics

	(1)		(2)		(3)		p-value (1)=(2)	p-value (1)=(3)	p-value (2)=(3)
	Comparison Schools		Adult Education Only		Adult Education + Monitoring				
<i>Panel A. Teacher Characteristics</i>	Mean	s.d	Mean	s.d.	Mean	s.d.			
Teacher Age			37.35	(8.67)	36.84	(9.37)			0.836
Teacher is female			0.33	(0.47)	0.34	(0.48)			0.816
Teacher is married			0.88	(0.33)	0.92	(0.27)			0.561
Teacher has some secondary education			0.35	(0.48)	0.39	(0.49)			0.569

Note: This table shows the difference in means between the different treatment groups. "Comparison" is defined as villages assigned to no adult education treatment in 2014 or 2015. "Adult education" is defined as those villages that were assigned to adult education without monitoring, whereas "Monitoring" is defined as villages that were assigned to adult education with monitoring. Standard deviations are shown in parentheses. Columns (4) and (5) show the coefficients and s.e. from a regression of each characteristic on the treatments and stratification fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 2. Reading Timed Z-Scores 2014-2015

	(1)	(2)	(3)	(4)	(5)	(6)
	Letters	Syllables	Words	Phrases	Comprehension	Composite Score
(1) Adult education	0.30*** (0.10)	0.23*** (0.09)	0.15* (0.08)	0.15* (0.08)	0.14* (0.08)	0.23*** (0.09)
(2) Adult education*monitor	0.18* (0.09)	0.19** (0.09)	0.14* (0.08)	0.10 (0.08)	0.07 (0.08)	0.16* (0.09)
Observations	3,481	3,482	3,482	3,478	3,482	3,482
R-squared	0.02	0.01	0.01	0.01	0.01	0.02
Total effect: Adult Education + Monitoring						
<i>p-value (Adult education + monitor=0)</i>	.00***	.00***	.00***	.01**	.05**	0.00***

Notes: This table presents the results from a regression of different reading outcomes on adult education (only), adult education plus monitoring and randomization fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 3. Math Z-Scores (Untimed), 2014-2015

	(1)	(2)	(3)	(4)	(5)	(6)
	Number Identification	Quantity Comparison	Addition and Subtraction (Simple)	Addition and Subtraction (Difficult)	Multiplication and Division	Composite Score
(1) Adult education	0.13*	-0.04	0.29***	0.23***	0.20**	0.20**
	(0.07)	(0.05)	(0.09)	(0.08)	(0.08)	(0.08)
(2) Adult education*monitor	0.11*	0.07*	0.12	0.09	0.09	0.12
	(0.06)	(0.04)	(0.08)	(0.08)	(0.07)	(0.07)
Observations	3,462	3,470	3,478	3,480	3,480	3,455
R-squared	0.01	0.00	0.02	0.01	0.01	0.02
Total effect: Adult Education + Monitoring <i>p-value (Adult education + monitor=0)</i>	.00***	0.21	.00***	.00***	.00***	.00***

Notes: This table presents the results from a regression of different math outcomes on adult education (only), adult education plus monitoring and randomization fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 4. Reading Timed Z-Scores by Year

	(1)	(2)	(3)	(4)	(5)	(5)
	Letters	Syllables	Words	Phrases	Comprehension	Composite Score
<i>Panel A: 2014</i>						
(1) Adult education	0.27*** (0.10)	0.23** (0.09)	0.13 (0.08)	0.14* (0.09)	0.14 (0.09)	0.22** (0.10)
(2) Adult education*monitor	0.20** (0.10)	0.24** (0.10)	0.15* (0.08)	0.12 (0.09)	0.06 (0.08)	0.19** (0.09)
Observations	1,760	1,760	1,760	1,759	1,760	1,760
R-squared	0.02	0.02	0.01	0.01	0.01	0.02
Total effect: Adult Education + Monitoring						
<i>p-value (Adult education + monitor=0)</i>	.00***	.00***	.00***	.02**	0.11	.00***
<i>Panel B: 2015</i>						
(1) Adult education	0.33*** (0.10)	0.24*** (0.09)	0.17** (0.08)	0.15* (0.08)	0.13* (0.08)	0.24*** (0.09)
(2) Adult education*monitor	0.15 (0.10)	0.13 (0.09)	0.12 (0.08)	0.08 (0.07)	0.09 (0.08)	0.12 (0.09)
Observations	1,721	1,722	1,722	1,719	1,722	1,722
R-squared	0.02	0.01	0.01	0.01	0.01	0.02
Total effect: Adult Education + Monitoring						
<i>p-value (Adult education + monitor=0)</i>	.00***	.00***	.00***	.02**	.03**	.00***
<i>Panel C: 2015 for Joint Monitoring</i>						
(1) Adult education	0.33*** (0.10)	0.24*** (0.09)	0.16** (0.08)	0.15* (0.08)	0.14* (0.08)	0.24*** (0.09)
(2) Adult education*monitor	0.18 (0.13)	0.14 (0.12)	0.16 (0.11)	0.13 (0.10)	0.16 (0.11)	0.16 (0.12)

Observations	1,334	1,335	1,335	1,333	1,335	1,320
R-squared	0.03	0.02	0.02	0.02	0.02	0.03
Total effect: Adult Education + Monitoring						
<i>p-value (Adult education + monitor=0)</i>	.00***	.00***	.02**	.02**	.03**	.00***

Notes: This table presents the results from a regression of different reading outcomes on adult education (only), adult education plus monitoring and randomization fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 5. Math Untimed Z-Scores by Year

	(1)	(2)	(3) Addition and Subtraction (Simple)	(4) Addition and Subtraction (Difficult)	(5) Multiplication and Division	(6) Composite Score
<i>Panel A: 2014</i>						
(1) Adult education	0.08 (0.07)	-0.03 (0.07)	0.22** (0.09)	0.16* (0.08)	0.17** (0.09)	0.15* (0.08)
(2) Adult education*monitor	0.14** (0.06)	-0.00 (0.06)	0.17** (0.08)	0.10 (0.08)	0.08 (0.08)	0.14* (0.07)
Observations	1,758	1,751	1,759	1,761	1,761	1,751
R-squared	0.01	0.01	0.02	0.01	0.01	0.01
Total effect: Adult Education + Monitoring <i>p-value (Adult education + monitor=0)</i>	.00***	0.89	.00***	.00***	.01***	.00***
<i>Panel B: 2015</i>						
(1) Adult education	0.18** (0.08)	-0.05 (0.06)	0.34*** (0.10)	0.28*** (0.09)	0.21** (0.10)	0.25*** (0.09)
(2) Adult education*monitor	0.08 (0.07)	0.16*** (0.05)	0.10 (0.08)	0.12 (0.09)	0.12 (0.08)	0.11 (0.08)
Observations	1,704	1,719	1,719	1,719	1,719	1,704
R-squared	0.01	0.01	0.03	0.02	0.02	0.02
Total effect: Adult Education + Monitoring <i>p-value (Adult education + monitor=0)</i>	0.00***	.01***	.00***	.00***	.00***	.00***
<i>Panel C: 2015 for Joint Monitoring</i>						

(1) Adult education	0.18**	-0.04	0.34***	0.29***	0.22**	0.25**
	(0.08)	(0.06)	(0.11)	(0.09)	(0.10)	(0.10)
(2) Adult education*monitor	0.12	0.21***	0.15	0.18	0.23**	0.17
	(0.10)	(0.06)	(0.11)	(0.12)	(0.10)	(0.11)
Observations	1,318	1,331	1,331	1,331	1,331	1,318
R-squared	0.02	0.02	0.04	0.03	0.03	0.02
Total effect: Adult Education + Monitoring						
<i>p-value (Adult education + monitor=0)</i>	.01**	.00***	.00***	.00***	.00***	.00***

Notes: This table presents the results from a regression of different reading outcomes on adult education (only), adult education plus monitoring and randomization fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 6. Heterogeneous Effects

	(1)	(2)
	Reading Composite Z-Score	Math Composite Z-Score
<i>Panel A: Effects by Region</i>		
(1) Adult education	0.17 (0.10)	0.16 (0.10)
(2) Adult education*monitor	0.17 (0.12)	0.15 (0.11)
(3) Adult education*Maradi	0.17 (0.18)	0.10 (0.16)
(3) Adult education*monitor*Maradi	-0.02 (0.17)	-0.04 (0.15)
(4) Maradi	-0.26 (0.25)	-0.05 (0.24)
Observations	3468	3,455
R-squared	0.02	0.02
<i>Panel B: Effects by Gender</i>		
(1) Adult education	0.42* (0.22)	0.31** (0.12)
(2) Adult education*monitor	0.26 (0.19)	0.05 (0.12)
(3) Adult education*female	-0.23 (0.23)	-0.12 (0.13)
(3) Adult education*monitor*female	-0.19 (0.19)	0.07 (0.12)
(4) Female	-0.76*** (0.18)	-1.21*** (0.10)
Observations	3,481	3,455
R-squared	0.14	0.25

Notes: This table presents the results from a regression of different outcomes on adult education (only), adult education plus monitoring, gender, the separate interaction terms and randomization fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 7. Heterogeneous Effects by Teacher Characteristics

	Reading Z-Scores				Math Z-Scores			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(1) Monitor	0.15 (0.10)	0.14 (0.10)	0.13 (0.10)	-0.10 (0.13)	0.11 (0.08)	0.11 (0.10)	0.06 (0.09)	-0.07 (0.13)
(2) Monitor*teacher is female	0.17 (0.19)				0.12 (0.16)			
(3) Female	0.04 (0.11)				0.07 (0.09)			
(4) Monitor*teacher has secondary education		0.08 (0.13)				0.02 (0.12)		
(5) Teacher has secondary education		0.03 (0.09)				-0.00 (0.07)		
(6) Monitor*teacher experience			0.03** (0.01)				0.03* (0.02)	
(7) Teacher experience			-0.05*** (0.01)				-0.04*** (0.02)	
(8) Monitor*Local Teacher (<= 5 km from village)				0.34** (0.16)				0.23 (0.14)
(9) Teacher is local				0.06 (0.10)				0.13 (0.10)
Number of observations	2,663	2,503	1,199	1,222	2,653	2,492	1,196	1,219
R-squared	0.01	0.01	0.02	0.02	0.02	0.01	0.02	0.03

Notes: This table presents the results from a regression of different reading and outcomes on monitoring, its interaction with different teacher characteristics (gender, education and experience), the teacher characteristics (not shown) and randomization fixed effects. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 8. Teacher Effort and Motivation

	Mean Non-Monitoring Village	Monitoring Village
	Mean (s.d.)	Coeff (s.e.)
<i>Panel A: Self-reported teacher attendance</i>		
(1) Stopped course (Yes/No)	0.53 (0.50)	-0.03 (0.06)
(2) Number of days stopped course	2.06 (4.12)	-1.30* (0.67)
<i>Panel B: Teacher Performance</i>		
Number of classes teachers taught (attendance lists)	19.25 (2.12)	-0.85 (0.92)
Teacher was replaced	0.24 (0.43)	-0.04 (0.07)
<i>Panel C: Teacher Motivation</i>		
Teacher kept an attendance log	0.28 (0.45)	0.18** (0.08)
Intrinsic motivation z-score	0 (1.00)	0.24* (0.13)
Perceived competence z-score	0 (1.00)	-0.02 (0.14)
Perceived pressure z-score	0 (1.00)	0.08 (0.12)
Perceived choice z-score	0 (1.00)	0.14 (0.12)
Number of observations		240

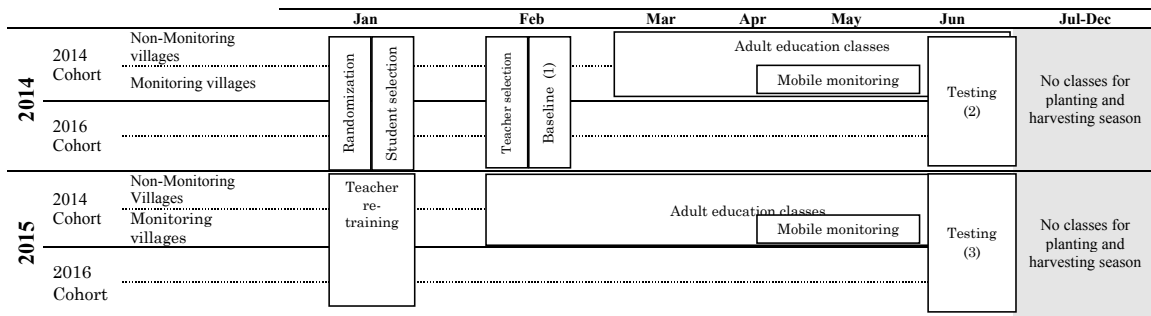
Notes: This table presents the results from a regression of teacher-level outcomes on a binary variable for monitoring, among the sample of adult education courses. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Table 9. Student Effort

	Adult Education Village	Adult Education*Monitor
	Mean (s.d.)	Coeff (s.e.)
<i>Panel A: Student Drop-Out of Course</i>		
Stopped course (Yes/No)	0.27 (0.44)	-0.02 (0.02)
Stopped course for personal choice (Yes/No)	0.11 (0.31)	-0.01 (0.02)
Length of time in course (months)	1.92 (1.23)	0.05 (0.08)
<i>Panel B: Learning Outcomes of Called Students (Compared with All Monitoring Students)</i>		
Reading z-score		0.58** (0.27)
Math z-score		0.24 (0.17)
Number of observations		1,773

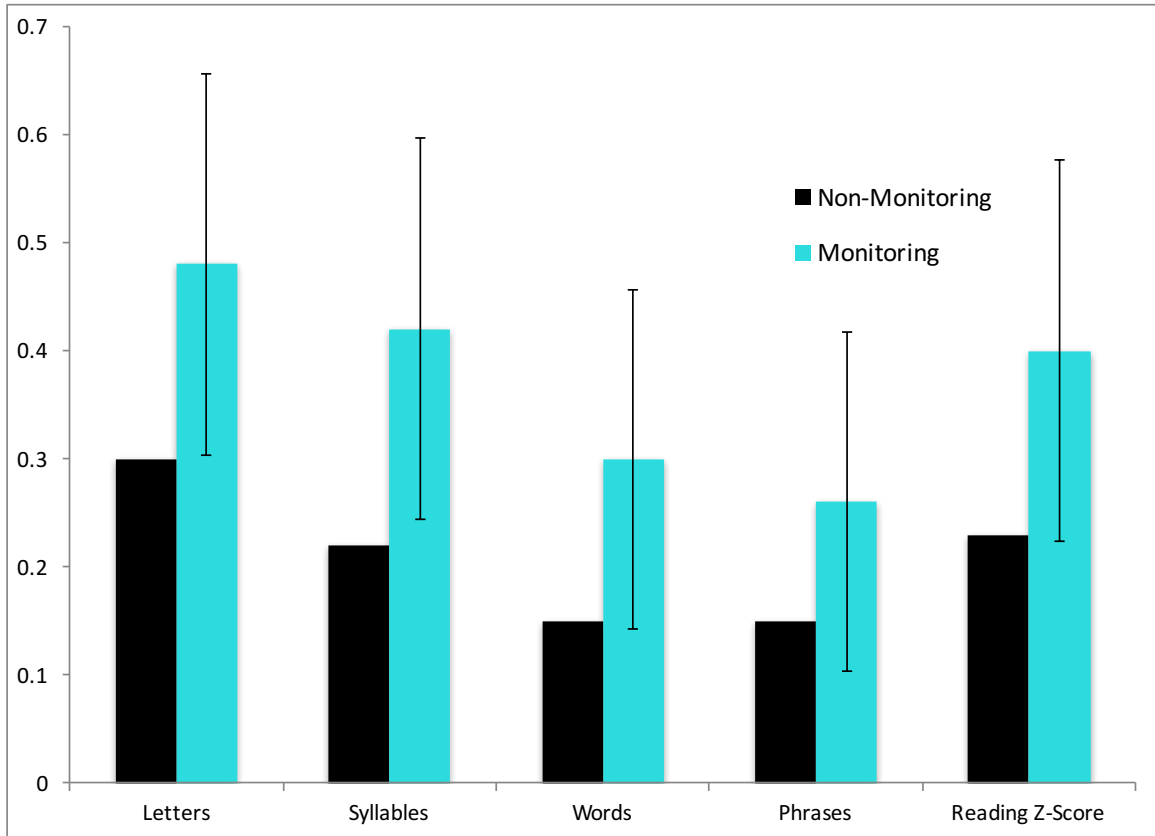
Notes: This table presents the results from a regression of student-level outcomes on a binary variable for monitoring, among the sample of adult education villages. Huber-White standard errors clustered at the village level are provided in parentheses. *** significant at the 1 percent level, ** significant at the 5 percent level, * significant at the 10 percent level.

Figure 1. Timeline of Activities



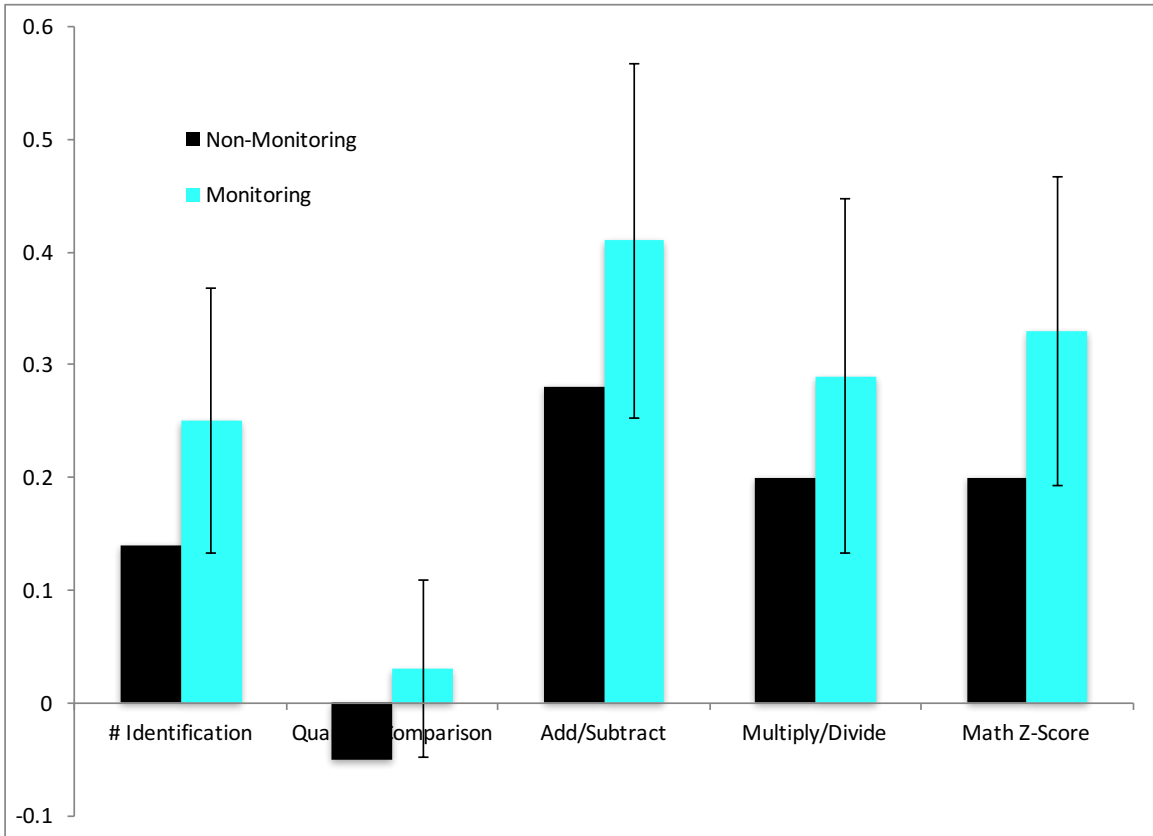
Note: Figure shows the timeline of activities for the different groups in our study. The 140 villages receiving adult education classes either did not receive extra monitoring attention (Non-monitoring villages) or received the mobile phone-based monitoring (Monitoring villages). The 2016 cohort is the group of 20 comparison villages, in which no adult education program was implemented in 2014, and which serve to estimate the impacts of the literacy program in concurrent research.

Figure 2A. Impact of the Monitoring Program over Both Years



Notes: This figure shows the mean timed reading z-scores of different reading tasks for students in monitoring and non-monitoring villages, controlling for stratification fixed effects. Timed reading scores are normalized according to contemporaneous reading scores in comparison villages. Standard errors are corrected for heteroskedasticity and clustered at the village level.

Figure 2B. Impact of Monitoring on Math Z-Scores over Both Years



Notes: This figure shows the mean math z-scores of different math tasks for students in monitoring and non-monitoring villages, controlling for stratification fixed effects. Math scores are normalized according to contemporaneous math scores in comparison villages. Standard errors are corrected for heteroskedasticity and clustered at the village level.

Figure 3. Costs of the Mobile Monitoring Intervention

