Learning without Teachers? Experimental Evidence on the Impact of a Mobile Phone Adult Literacy Program

Christopher Ksoll*, Jenny C. Aker*, Danielle Miller, Karla Perez-Mendoza, Susan L. Smalley[1]

August 2023

Abstract: Could a simple technology help low-literate adults learn to read without teachers? We conduct a randomized control trial testing the impact of a Spanish literacy program implemented via simple mobile phones in Los Angeles. The program almost doubled adult learners' reading scores as compared with the control, as well as improved their self-esteem. Learners used the platform at all times of day and for different durations, suggesting the importance of flexibility. Our findings have policy relevance regarding the role of simple technology in providing learning support to illiterate and low-literate adults, especially in contexts where it is difficult to find teachers or hold classroom-based learning.

(JEL code: I21)

Keywords: Adult Education; Information and Communications Technology; E-Learning

[1]Christopher Ksoll, University of Sherbrooke, Sherbrooke, Canada. christopher.ksoll@usherbrooke.ca. Corresponding author. Jenny C. Aker, The Fletcher School and Department of Economics, Tufts University, 160 Packard Avenue, Medford, MA 02155; Jenny.Aker@tufts.edu.. Danielle Miller, Dornsife College of Letters, Arts and Sciences, University of Southern California; KeenanMi@usc.edu. Karla C. Perez-Mendoza, Graduate School of Education, and Information Studies (GSE&IS), University of California Los Angeles; kcperez77@ucla.edu. Susan L. Smalley, Department of Psychiatry and Behavioral Science, UCLA; sue@suesmalley.com. We thank Michael Clemens, Jose Galdo, Louis-Philippe Morin, Concepcion Valadez, Bruce Wydick, as well as seminar participants at the University of San Francisco, Tufts University and Mathematica Policy Research for helpful comments. We thank Ariana Cornei, Stephanie Del Mel, and Wei Yang for excellent research assistance. Smalley provided funding for the data-collection. Ksoll acknowledges funding from Canadian Social Sciences and Humanities Research Council (SSHRC) Insight Grant No 435-2013-1913.

Despite steady improvements in literacy rates over the past fifty years, 14% of adults worldwide are unable to read and write in any language (UNESCO 2020). These statistics are not only confined to low-income countries: In the US, 54 percent of adults lack literacy proficiency, reading below a sixth-grade level (US Department of Education).[2] While adult education campaigns have often been proposed to meet the educational needs of illiterate adults, studies on their impacts show limited educational gains and rapid skills depreciation (Abadzi 1994, 2003). These poor learning outcomes may be due, in part, to high student drop-out, as there are often significant opportunity costs and uncertain returns associated with traditional adult education programs (Aker et al 2023).

Given the widespread growth of mobile phone coverage over the past decade, a number of technology-based adult education programs have been developed. Most of these programs require smart phones or internet access, which are not easily accessible to poorer populations or in remote areas. Furthermore, such programs are often complements to teachers and thus are heavily dependent upon teacher availability and quality, which is not readily available in low-resource contexts.

[2] A person is considered illiterate if they are unable to read and understand basic prose in any language. A person is considered functionally literate if they can "engage in all those activities in which literacy is required for effective functioning of his group and community and also for enabling him to continue to use reading, writing and calculation for his own and the community's development." (UNESCO 2005)

In this paper, we report the results from an experiment in Los Angeles, whereby a mobile phone-based adult education program in Spanish was offered to low-literate Hispanic adults (*Cell-Ed*). The *Cell-Ed* learning curriculum was provided entirely via a series of voice and SMS-based operations on the mobile phone and did not require teacher instruction or in-situ learning. After a detailed recruitment and screening process, whereby participants' baseline literacy levels were assessed, seventy adults were randomly assigned to the treatment (*Cell-Ed*) or control group. We find that the program substantially improved learning outcomes: The increase in *Cell-Ed* participants' reading skills after four months was equivalent to the reading skills children acquire after approximately two years of schooling. We posit that these increases may be due in part to the flexibility of the curriculum, as learners opted to learn at all times of the day and for short durations.[3]

A key claim of many adult education programs is that they empower learners by providing them with the knowledge and skills necessary to acquire new labor market opportunities and live more independent lives. While we do not have information on labor market outcomes, we investigate whether the *Cell-Ed* program affected psychological measures,

[3]One potential limitation of our study is the relatively small sample size. The power calculations that informed the sample size were based on effect sizes from a previous adult education program in Niger, which suggest that our initial sample for this experiment was not underpowered (Aker et al 2012). We also show that the magnitude of the impact of the control group after phase-in is similar to that of the treatment group, which suggests that the treatment effects are not an artifact of the small sample size. Nevertheless, the sample size does affect the generalizability of our findings.

as measured by the Rosenberg self-esteem score (RSES) and the General Self-Efficacy Scale (GSES). Using real-time data on learning progress on the platform and weekly measures of self-esteem, we also provide insights into the dynamic relationship between learning progress and self-esteem.

A potential threat to the internal validity of our findings is differential attrition between the treatment and control groups. To deal with non-random attrition, we implement a non-parametric bounding approach following Lee (2009), which bounds the treatment effect away from zero, and the lower bounds are large. We are able to conclude that the *Cell-Ed* program is highly effective in improving reading levels, even when using very robust approaches to account for attrition.

Our results contribute to the growing debate on the effectiveness of digitally-assisted learning in other contexts. A number of papers find that computers and laptops have mixed results on primary school students' learning outcomes, from no effects (Barrera-Osario and Linden 2009, Fairlie and Robinson 2013) to positive effects (Banerjee et al 2007, Deshpande et al 2017, Barrow, Markman and Rouse 2009). Focusing on simple digital technologies (mobile phones), Aker et al (2012) find that learning how to use mobile phones as part of an adult education class in Niger increased test scores by .19-.25 s.d., whereas Angrist et al (2022) find that simple phone calls and SMS to parents increased children's test scores by .12 s.d during the COVID-19 crisis. While most of these papers use digital technologies as a complement to in-person learning, our experiment provides evidence that a simple, self-directed educational device can lead to substantial improvements in learning.

Our results also contribute to the small literature on the impacts of adult education programs on learning outcomes. Despite decades of experience with adult education programs, as well as significant public spending in this area, there is a relative paucity of evidence (Aker et 2023). Existing studies show that such programs can improve learning, but that the magnitudes are somewhat small (Banerji et al 2017, Deshpande et al 2017, Aker and Ksoll 2019). By contrast, our results show large learning gains in reading over a short period of time.

Finally, our paper speaks to the relationship between education and socio-emotional well-being. While there is significant literature measuring the impact of education programs on education and labor markets outcomes, health and fertility (Case 2005), empowerment and economic development (Duflo 2012, Doepke and Tertilt 2014) or the impact of different interventions on intra-household decision-making (Ashraf 2009, Ashraf et al 2010), there is more limited evidence of the link between education and empowerment. More recently, Carlana and La Ferrara (2021) found that an online tutoring program implemented during the COVID-19 pandemic significantly improved middle school students' psycho-social well-being. Yet other studies of adult education programs have found no significant evidence of impacts of adult education programs on women's empowerment (Banerji et al 2017, Deshpande et al 2023). Our paper contributes to this literature, not only by finding a positive effect, but also by measuring the dynamic relationship between education and psycho-social outcomes.

The remainder of the paper is organized as follows. Section I provides background on the setting and the research design. Section II describes the different datasets and Section III outlines the estimation strategy. Section IV discusses the results, whereas Section V looks at the potential learning mechanisms. Section VI concludes.

I. Research Setting and Design

While the United States is one of the wealthiest countries in the world, it is estimated that 1 in 7 U.S. adults are functionally illiterate, defined as being unable to read and understand basic prose (National Center for Education Statistics 2003). The Hispanic population accounts for a disproportionate share of the functionally illiterate, with approximately 44 percent of Hispanic adults having literacy levels considered to be at or "below basic" level. Hispanic populations also represent 63% of the "least literate" adults in the U.S., defined as those who fail the simplest screener questions.[4]

[4]The National Assessment of Adult Literacy is a nationwide representative survey on literacy in the US. The assessment first asks respondents 7 very simple literacy questions. If respondents fail this initial screening (about 3% did), they participate in a survey for the "least literate adults". 63% of the respondents in this category are Hispanics. Of the three percent, only 57% could read any letters and 46% could read any words.

Greater Los Angeles has the highest rate of "undereducated adults" in any major U.S. metropolitan area (Literacy Network of Greater LA and United Way 2004).[5] Similar to national figures, a disproportionate number of the low-literate and illiterate populations are immigrants, primarily Spanish-speaking immigrants who are unable to read or write Spanish or English.

In an effort to address this issue, the U.S. Adult Education and Family Literacy Act provides approximately US$ 600 million per year in funding for states to implement adult education programs. A majority of these adult education programs are taught using teacher-based personal instruction, often in adult schools or community centers. Despite these efforts, adult education programs reach fewer than 16 percent of the low-literacy population in Los Angeles, and drop-out rates are well over 50 percent (Literacy Network of Greater LA and United Way 2004).

A. Cell-Ed Intervention

The widespread growth of mobile phone coverage has the potential to facilitate skills acquisition of illiterate adults, as well as increase the scope and scale of such programs. While a number of technology-based adult education programs have been developed, most require

[5]The Literacy Network of Greater LA and United Way (2004) define literacy and numeracy levels slightly differently (National Center for Education Statistics 2003). In their definition, individuals with a literacy "Level 1" are unable to engage in basic daily tasks, such as find an intersection on a map. Approximately 32 percent of the population in Greater LA (or 2.3 million people) falls into this category, as compared with 20 percent at the national level.

internet access or smart phones, which are often not available among poorer populations in Los Angeles. *Cell-Ed* is a platform that provides basic educational instruction via simple mobile phones. The platform uses voice and SMS to deliver 437 adult education lessons (called "micro-modules") to learners. Each micro-module consists of three components: 1) an *audio lesson* on a particular concept (vowels, consonants, words), varying from 1-3 minutes in length, is introduced when the learner calls a designated number; 2) an SMS message reinforcing the voice lesson is sent to the participant; and 3) an SMS question is sent to the participant asking them about the lesson that they recently learned, and the participant must text a response. A correct response triggers the beginning of the next micro-module, whereas an incorrect response leads to a repetition of the same micro-module until the user succeeds. To activate the program and each micro-module, participants call the *Cell-Ed* phone number from their own mobile phone.[6] Students can access the program 24 hours a day, seven days a week, allowing them to learn when, where and how they wish.

The *Cell-Ed* curriculum for this study was based upon a traditional Spanish adult education program (LEAMOS!) developed by the Centro Latino in Los Angeles. The traditional LEAMOS! curriculum is comprised of 43 lessons and teaches simple Spanish letter and word

[6]If *Cell-Ed* participants did not own a mobile phone, they were provided with a simple phone as part of the program. All participants (regardless of mobile phone ownership) were also provided with an unlimited voice and SMS plan if they did not already have one.

recognition, as well as reading and writing skills. The curriculum is typically taught by an in-classroom teacher and takes approximately 150 hours to complete over a four-month period, or 9 hours per week. The 43 LEAMOS! lessons were adapted into 437 micro-modules for the *Cell-Ed* platform, with each micro-module including recorded audio instructions, SMS messages and queries for interactive testing. In addition to the micro-modules, pre-recorded audio messages were sent to each learner to offer encouragement at various points in the process.

B. Experimental Design

Prior to the introduction of the program, *Cell-Ed* partnered with two schools and five community resource centers with large Hispanic populations in the Los Angeles area. A variety of recruitment methods were used to recruit potential participants, including informational flyers, door-to-door visits by parent volunteers and presentations at community meetings, school fairs and school events. Using these recruitment methods, we identified 250 individuals and conducted an initial screening process via phone. The screening process collected information on individuals' socio-demographic characteristics, reading ability, mobile phone access and eyesight. Participants were excluded from further participation if they were over the age of 80 years, could read sentences, or required eyeglasses and did not own them, thereby reducing the sample size to 124 participants.[7]

The 124 participants were asked to attend an additional in-person baseline evaluation between March and July 2012, of which 89 attended. Those who attended were provided with an ID number, which were divided into strata and randomly assigned to either the treatment (the *Cell-Ed* program) or the control group (to be phased into the *Cell-Ed* program approximately 3-4 months later) with equal probability.[8] The random assignment took place before the evaluation, and enumerators were not informed of participants' treatment or control status.

The baseline evaluation participants were asked to complete a battery of assessments, including the Woodcock-Muñoz Achievement Battery. Participants who scored sufficiently low on the reading test were considered eligible for participation in the study.[9] This exclusion further reduced the eligible sample to 70 participants. After finishing the screening process, participants were informed of their treatment status via a previously sealed letter (linked to

[7]This initial screening process implies that our reduced sample was less literate than the initial sample of 250 individuals, but had access to eyeglasses, if necessary.

[8]In order to ensure that balance between treatment and controls was achieved for earlier and later potential participants, we stratified the random assignment by pre-screening ID in the following strata: 1-22,23-46,47-70,70-100,100+. Within strata the ex-ante probability of selection into treatment was 50 percent. We control for strata fixed effects in all OLS as the ex-post probability differed due to ineligible applicants.

[9]Participants who attained a level of "having basic literacy skills" on more than one of three reading sub-tests were excluded. This exclusion criteria was put in place for two reasons: First, to reduce the variance of baseline scores, and second, to understand whether the least literate would find the technology manageable.

their initial ID) that indicated whether they had been assigned to the treatment or control condition. This letter was read to them.[10]

The treatment thereby followed a randomized phase-in design. After the initial baseline evaluation, those in the treatment group were asked to return for a second evaluation approximately 4.5 months later, in order to ensure that they had sufficient time to complete the *Cell-Ed* program.[11] The control group was eligible to start the *Cell-Ed* program approximately three months after the baseline, after having completed a midline survey. A timeline of the implementation and data collection activities is provided in Figure 1.

For each survey round, participants were provided with compensation in the form of a US$50 gift card to a local supermarket chain. In addition, there was a US$100 incentive for completing the program in 4.5 months. Thus, a treatment participant that completed all survey rounds and finished the program in 4.5 months would receive US$200, whereas a control participant who completed all survey rounds and finish the program in 4.5 months (after being

[10]As the control group was informed that they would receive the treatment at a later time, this could have changed their behavior between the first and second survey rounds (the John Henry effect). If the control group exerted less effort in learning than they would have otherwise, our results could potentially provide an upper bound for the treatment effect. If, however, this letter encouraged the control group participants to exert greater effort in learning during the first round, then our results would provide a lower bound on the treatment effect. As reading skills are difficult to acquire without some type of external support, we do not think that the John Henry effect is a primary concern in this context.
[11]Early piloting of the program suggested that most participants could complete the program in 4.5 months. This time period was also similar to the time frame for most Leamos! learners.

phased in) would receive US$ 250.[12] Overall, the screening process suggests that our sample is less educated than the broader population of Spanish-speaking immigrants in Los Angeles, as they did not perform well on initial reading tests.

II. Data

This paper relies upon four primary datasets. First, we administered comprehensive reading tests to measure the impact of *Cell-Ed* on learning outcomes. Second, we collected data on student and household characteristics, as well as some qualitative data on their experiences in Los Angeles. Third, we conducted short weekly phone calls to both treatment and control participants. Fourth, we collected real-time *Cell-Ed* usage data, providing user-specific information on how much time was spent on the program, as well as performance statistics (i.e., whether the student failed or passed a specific module).

A. Test Score Data

Reading tests were administered to all participants during the baseline survey, providing a sample of 89 students (of which 70 were retained for the experiment). We administered a second round of reading tests for both the treatment and control groups approximately four months after the baseline, with a third round of follow-up tests for the control group only 4.5

[12]As the intervention bundled learning incentives with the learning technology, we are unable measure the impact of the learning incentives separately from that of the technology.

months after they started the *Cell-Ed* program. The second set of test scores allows us to measure the causal impact of the *Cell-Ed* program on learning, as the control group had not yet started the program. The third set of test scores for the control group allows us to assess whether the impacts were similar in the control group as they were in the initial treatment group.[13]

The test score data were based upon the Spanish language equivalent of the widely used Woodcock-Johnson battery of literacy tests, the *Woodcock-Muñoz III Language Survey (WMLS-III).* The typical survey includes a screener and a more-comprehensive seven-test battery, although our survey did not include the former.[14] We administered the cognitive battery of tests, which provides a composite verbal IQ score, as well as the reading achievement battery of tests, which were used to calculate two composite scores: the *basic reading score*, which comprises two sub-tests (letter-word identification and word attack); and the *broad reading score*, which comprises three sub-tests (letter-word identification, reading fluency and

[13]For treatment individuals who completed the program in less than 4.5 months, the first follow-up survey was done as soon as possible after completion to minimize survey non-response. Typically, follow-up surveys would be implemented at the same time for both the treatment and control groups. In our case, the change over the course of 3 months in reading ability in the absence of treatment is zero, implying that this would be the case 4.5 months later as well, unless the comparison group learned how to read otherwise. Testing the treatment group when they finished the program (with a maximum date of 4.5 months) alters the interpretation of the estimate: It is not the impact on learning after 4.5 months, but the impact of *Cell-Ed* at the individually defined end of the program, censored after 4.5 months. Some learners continued after the 4.5 months, and this learning is not captured in our tests.

[14] The typical seven tests involve picture vocabulary, verbal analogies, letter-word identification, dictation, understanding directions, story recall and passage comprehension. Testing times range from approx. 25 minutes (screener) to 55 minutes (complete battery).

passage comprehension). The basic reading score covers literacy skills as narrowly defined, namely reading decoding and phonetic coding (identifying letters and building them up to words) (McGrew et al. 2007). The broad reading score can be interpreted as a measure of more advanced reading skills, as it contains dimensions such as fluency and comprehension. Each composite score is calculated as an "age equivalent", which indicates the typical age of persons in the population who obtain a given score. For example, if a student's performance on the test of reading comprehension is equal to an age equivalent of 8.5 years, this means that his or her obtained *raw score* is equivalent to the predicted average for 8-year, 6-month old children in the norm group. We use the age equivalent throughout our paper since it has an intuitive interpretation.[15] As Spanish is a phonetic language, where each letter maps into one sound, becoming literate may be simpler than in English, where the mapping between letters and sounds is not unique (Abadzi 2013).

B. Socio-Demographic Characteristics and Measures of Empowerment

The second primary dataset includes information on student and household characteristics, collected at the same time as the test score data. The survey collected detailed information on students' demographics (age, gender, birthplace), education, employment,

[15] We also provide results with z-scores based on the norming population of the tests.

household size, mobile phone ownership and usage. We use these data primarily for balance checks and additional controls.

We also collected data on two measures of psychological dimensions of well-being: self-esteem and self-efficacy, as measured by the Rosenberg self-esteem scale and the general self-efficacy score. The Rosenberg self-esteem scale (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 maximum is 30). The General Self-Efficacy Scale (GSES) (Schwarzer and Jerusalem 1995) is also a 10-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. 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 *Cell-Ed* program on participants' perceptions of empowerment.

C. Weekly phone calls

In addition to the in-person surveys, we contacted all participants via weekly phone calls. Treatment participants were asked whether anyone assisted them with *Cell-Ed* that week, whether they had technical difficulties, what they liked and disliked about the curriculum and whether there were any impediments to their studies. They were also asked about mobile phone usage and self-esteem. As the phone call might be an intervention in itself, phone calls were also made to the control group. Control participants were asked the same questions on mobile phone usage and self-esteem, but not about the *Cell-Ed* program.

D. Cell-Ed Usage Data

The final dataset comprises records from the *Cell-Ed* platform, which logs every interaction between the student and the platform. The log includes the date and time the participant called in, the mini-module they accessed, the timing and content of SMS messages received and sent, and participants' performance on test questions for each micro-module. Using these data, we calculate a number of statistics, including the number of days between different modules; how the student performed on each module (including the number of attempts to obtain a correct response); how quickly each student completed the program; and whether the student stopped calling (and hence dropped out of) the *Cell-Ed* program. We use these data to provide insights into the mechanisms through which the *Cell-Ed* program affected learning outcomes.

E. Pre-Program Balance

Table 1 provides an overview of student characteristics by treatment status. The evidence suggests that the randomization was successful in creating comparable groups along observable dimensions. Differences in students' characteristics between the treatment and control group before the program started are small and not statistically significant (Table 1, Panel A). Participants were 47 years old on average and a majority of participants were women.

While 51% percent of participants had some schooling, the average duration of schooling was quite low: 1.22 years. 51 percent of households in the sample owned a mobile phone, broadly in line with other studies on mobile phone ownership among Hispanic populations with lower levels of education.[16] Among all of these variables, only the difference in mobile phone ownership is statistically significant at the 10 percent level, with a higher mean in the treatment group as compared with the control group. As mobile phone ownership could be positively correlated with the treatment and our outcomes of interest, we control for baseline levels of mobile phone ownership as a robustness check.

Panel B shows the means of key outcome variables (broad and basic reading age equivalent, the Rosenberg score and the self-efficacy score) by treatment status. During the baseline, participants had mean basic and broad reading scores of 6, suggesting that their reading levels were equivalent to those of a six-year-old. There was no statistically significant difference between the two groups. While the results of the Generalized Self-Efficacy score were similar across treatment and control groups, participants in the treatment group had slightly higher baseline levels of self-esteem, with a statistically significant difference at the 10 percent

---

[16]A survey by the Pew Trust estimated that mobile phone ownership among Hispanic households with lower income and educational levels ranged from 56-77%. http://www.pewhispanic.org/2013/03/07/closing-the-digital-divide-latinos-and-technology-adoption/.

level. As this measure is potentially correlated with treatment and the outcomes of interest, we control for baseline levels of self-esteem in the robustness checks as well.

A concern with the baseline balance tests is that the absence of any statistically significant differences may be due to our limited number of observations, hence increasing the size of our standard errors. As a result, we implement specifications that account for such differences, as outlined below.

III. Estimation Strategy

All of the 70 participants who were initially assigned to the *Cell-Ed* platform were provided with access to the platform and used it (even if they dropped out at a later time). In addition, none of the initial control group participants accessed the *Cell-Ed* platform during the first phase of the experiment.[17] As a result, there was no imperfect compliance: the treatment assignment variable is equivalent to the treatment participation variable, and our intention to treat (ITT) effect is equivalent to the average treatment effect on the treated (ATT).

Let $test_i$ be the composite basic or broad reading test score attained by student i after the program. $Cell\text{-}Ed_i$ is an indicator variable for whether individual i is assigned to the *Cell-Ed* intervention ($Cell\text{-}Ed$=1) or the control ($Cell\text{-}Ed$=0) in the first period. θ_R are fixed effects

[17] The minimum time between the first and last contact with the platform among all treatment observations was 42 days, even including those who eventually dropped out. When the control group was phased into the treatment, the minimum number of days between first and last interactions on the platform was 21 days.

that indicate the randomization strata, based on the ID number. $\mathbf{X}'_{i0}$ is a vector of student-level baseline covariates, primarily gender, age and a proxy for IQ. Ignoring attrition, we first estimate the following specification:

$$test_i = \beta_0 + \beta_1 CellEd_i + \mathbf{X}'_{i0} + \theta_R + \varepsilon_i \quad (1)$$

The coefficient of interest is β_1, which captures the causal impact of the *Cell-Ed* program on learning outcomes. We also modify this specification to include the baseline outcome variable (a value-added specification), covariates that were statistically significant at baseline (mobile phone ownership and self-esteem) and a difference-in-differences (DD) estimation strategy.

While randomization would normally imply that β_1 has a causal interpretation, this is only the case if we assume that attriters (non-respondents) would have progressed in their learning as much as those who continued with the program.[18] An unbiased estimate of the ATT can only be estimated under the assumption that non-response is random. We discuss drop-out and non-response in more detail below.

[18]An alternative assumption would be if attrition were equally high in treatment and control groups, and we assume that treatment did not affect dropout for any survey participant. Then the estimated parameter would have the interpretation of the average treatment effect on the treated. In our survey there is a statistical difference in attrition, so this alternative identification assumption is untenable.

A. Non-Response and Drop-outs

Table 2A shows the rates of survey non-response by treatment status and across all survey rounds. While definitions of survey non-response (attrition) and dropout can differ, in our case, dropping out of the *Cell-Ed* program was strongly correlated with survey non-response, and so we treat them interchangeably. In our context, drop-outs comprise: 1) treated participants who could not make it past the first ten lessons;[19] 2) treated participants who made it past the first ten lessons but refused a follow-up interview; or 3) control participants who did not participate in follow-up interviews, either because they refused a follow-up interview or dropped out after the first ten lessons.

Overall, survey non-response was 17 percent during the first follow-up survey, with significant differences by treatment arm: the rate of attrition was 25 percent in the treatment group and 8.8 percent in the control group, with a statistically significant difference between the two. The rate of dropout was higher during the second survey round (once the control group started the program), where attrition was 32 percent. While these rates of attrition are high, we note that they are lower than what is typically observed in most adult education programs. The Literacy Network of Greater LA and the United Way (2004) estimate that 50

[19]Follow-up data were not collected from this group of non-respondents.

percent adult education students drop out within the first three weeks of such programs, with similar rates reported by Romain and Armstrong (1987).[20]

The difference in means for the baseline characteristics of respondents and non-respondents are presented in Table A1. Overall, non-respondents were largely similar along observable dimensions as compared with respondents, although part of this could be explained by our limited power to detect a statistically significant difference.[21] Non-respondents were more likely to have attended formal school and scored slightly higher on the baseline self-efficacy score (Table A1). These correlations are confirmed when looking at the determinants of dropout: attriters have a higher self-efficacy score than non-attriters, and were also more likely to have attended school (Table A2). In addition to these covariates, Table A2 also suggests that older participants and females were more likely to attrit.[22]

Table 2B presents more detailed information on the correlates between attrition and learning, based upon the *Cell-Ed* usage data. Whereas all students spent some time on the *Cell-Ed* platform, there are marked differences in the duration of time spent by respondents and

[20] Literacy agencies covering almost 600,000 learners responded to their survey, of which only 55 percent tracked information on learner retention.
[21] Looking at a broader range of covariates, initial treatment non-respondents differed at the 10% level from respondents on 8 variables out of 68. There were no statistically significant differences at the 5% level.
[22] An additional potential threat to the validity of our findings is spillovers. We cannot rule out the possibility that some treatment and control group participants knew each other before the program. In this case, however, our treatment effect estimates would represent a lower bound on the treatment effect in the absence of spillovers.

non-respondents (i.e., those who eventually dropped out). Focusing on the treatment group (Panel A), respondents interacted with the platform over the course of four months, as compared with three months for non-respondents. Respondents in the treatment group spent approximately 50 hours on the platform, whereas treatment group non-respondents only spent 8 hours on the platform. Finally, treatment respondents completed 402 modules (out of 437), about 90 percent of the curriculum, whereas treatment non-respondents completed only 7 percent of the curriculum. Once the control group phased into the *Cell-Ed* platform, the difference in usage patterns between respondents and non–respondents follows a similar pattern (Panel B).

In order to investigate whether the large results we find in the OLS regressions are driven by differential attrition, we implement two sets of non-parametric bounds following Manski (1997) and Lee (2009).

The monotone treatment response assumption (Manski, 1997) can be invoked when being treated cannot reduce a respondent's outcome. In our case all treatment respondents enjoyed some access to the platform and also spent some time on the platform. It is hard to see how the treatment could have lowered their literacy skills. A natural lower bound for their literacy skills are thus their own baseline test results. The monotone treatment response assumption is silent about the three control group observations that were non-respondents. We proceed in the same way, that is to impute their baseline scores, for two reasons: first, because for control group respondents as a group, there is no statistically significant difference between baseline and the first round (which suggests using the baseline is appropriate). Second, we want to have a consistent approach with the treatment non-respondents.

Lee's (2007) method acknowledges that we do not know with certainty what the outcomes are of participants who are not observed, but proposes that the outcomes of participants in the group with more attrition (also called group with excess attrition) can be bounded using the largest and smallest values in the group with lower attrition. Contrary to Manski (1990) bounds, this is not the single largest and smallest value, but the largest values that correspond to the proportion of excess attrition. In our case there are more treatment non-respondents than treatment respondents. Under the assumption that control group observations would also have dropped out if they had been treated, an extreme assumption on what the outcomes of the missing treatment respondents are if they had not received treatment are the best and worst outcomes in the control group. To estimate the maximum and minimum treatment effect, respectively, we drop control observations with the largest and smallest values up to the point that the proportion of observations included in the "trimmed" comparison group are the same as in the treatment group. Then we proceed with the analysis comparing treatment and trimmed control groups. The estimated values are bounds between which we are very confident that the true value would lie.

Applied to our context, the methodology is applied as follows: The treatment group is divided into two groups: respondents (with a sample size (N) of 27) and non-respondents (N=9). The control group also comprises the respondents (N=31) and the non-respondents whom we call "Always non-respondents".[23]

Lee bounds assume monotonicity, in other words, that the likelihood of non-response is monotonically related to receiving the treatment (Lee 2009). This is equivalent to stating that receiving the treatment either makes *all* observations (weakly) *more* likely to respond to the survey, or *all* observations (weakly) *less* likely to respond.[24] This assumption rules out the possibility that the treatment may affect different sub-groups differently, such as increasing the likelihood of response for one specific sub-group while decreasing the likelihood for another. This also implies that the non-respondents in the group with lower non-response would not have responded if their treatment status were changed. When attrition is higher in the treatment group, as is the case in our experimental set-up, the monotonicity assumption implies that the control observations who did not respond in the second survey would also not have responded had they immediately received treatment (Figure 2).[25]

[23] There is no ranking of outcomes associated with this graph, as observations in the phase-in non-respondents could well have expected outcomes that are higher than the lower bound of the phase-in respondents.
[24] We write "weakly" because it is possible that the treatment leaves some observations' likelihood unchanged.
[25] To be exact, the term "always non-respondent" thus already embodies the assumption we are making.

Under these assumptions, we first estimate the upper and lower bounds by assuming the "best" and "worst" case scenarios.[26] In our case, the treatment group has an attrition rate of 25 percent and the control group has an attrition rate of 8.8 percent, so the difference in drop-out is 16.2 percentage points. To estimate the *lower bound*, we trim the lowest 17.7 percent of the *observations* from the control group.[27] The lower bound is thus equivalent to assuming that the outcomes of additional treatment non-respondents correspond to those control group observations who make up the bottom end of the control group distribution: We compare the observed observations in the treatment group (who make up 75% of all the treatment group observations) to the 82.3% "best" observed observations in the control group (who make up 75% of all control group observations) (Figure A1). As is usual, the confidence intervals around this estimated *lower bound* are presented. We present bootstrapped confidence intervals around all bounds. To create the *upper bound*, we trim the highest 17.7 percent of observations from the control group (equivalent to 16.2 % unconditionally). The upper bound is equivalent to assuming that the additional non-respondents are the top learners in the control group, and is constructed by comparing the observed treatment group outcomes to the 82.3% worst control group observations (who make up 75% of all control group observations) (Figure A2).

---

[26]This assumes that the additional percentiles of non-respondents in the treatment group correspond to, respectively, the best and worst respective percentiles of observed outcomes in the control group (Lee 2009).

[27] Note that we calculate the proportion of observed control group observations that is trimmed as the difference in attrition divided by the proportion of control group observations that are observed is 17.7=16.2/(100-8.8). Because of indivisibilities, the exact trimming proportion is 19.35 in the empirical application.

IV. Results

Figure 3 depicts the mean reading scores for the treatment and control group respondents for the first survey round. Before the program, both groups had basic reading scores corresponding to an age equivalent of 6.5 years, without a statistically significant difference between the two (Table A3). Immediately after the program, *Cell-Ed* participants had basic reading skills equivalent to an 11.5 year-old child, with a statistically significant difference at the 1 percent level. Thus, *Cell-Ed* participants moved from a first-grade to a sixth-grade reading level, whereas the comparison group showed no improvements in reading during this time. Using broad reading scores, *Cell-Ed* participants' scores were 2-3 years higher than those in the control group, with a statistically significant difference at the 1 percent level. These results are similar for the control group once they phased into the program (Table A3), suggesting that the treatment effect was stable over time.[28] Figures A1 and A2 show that the treatment group's reading score distribution shifted sharply to the right.

A. Impact of Cell-Ed Program on Learning Outcomes

Table 3 presents the results of equation (1). Using the simplest specification, the *Cell-Ed* program increased students' basic reading scores by 5.65 years, with a statistically significant

[28]As there are only small baseline differences in learning outcomes between the treatment and control group, and no significant changes in the control group's learning outcomes between the baseline and first follow-up survey, any difference in impacts between these two groups would point to technical difficulties with the platform, spillover effects or learning on the part of the implementers. Given that the program is highly standardized through the platform, the absence of a difference is not too surprising.

effect at the 1 percent level (Column 1). These effects are robust to the inclusion of baseline covariates, namely age, gender and IQ (Column 2), a value-added specification (Column 3), a difference-in-differences specification (Column 4) and controlling for individual fixed effects (Column 5). The effects are also robust to the inclusion of the covariates that were different at baseline, as well as using normalized test scores as the dependent variable (Table A4). Overall, these results suggest that participating in the *Cell-Ed* program increased students' basic reading scores by 4.76 - 5.84 years within a four-month period, with a statistically significant effect at the 1 percent level.

The estimates are smaller in magnitude when using broad reading scores, which capture a broader set of reading skills: the *Cell-Ed* program increased broad reading scores by 2.58 years on average when using the simplest specification, with a statistically significant effect at the 1 percent level (Column 6). These effects are similar when including baseline covariates (Column 7), a value-added specification (Column 8), a DD specification (Column 9) and controlling for individual fixed effects (Column 10). The results are also robust to the inclusion of the baseline mobile ownership and self-esteem scores, as well as using normalized test scores as the dependent variable (Table A4).[29] The smaller impacts of the program on broad reading scores relative to

[29]Table A3 presents evidence that the control group benefits as much from the treatment when they are phased-in. Column 4 tests for difference in outcomes between treated treatment group respondents and treated control group phase-in respondents and cannot reject that they are equal.

basic reading scores is not surprising, as the *Cell-Ed* program focused on basic reading skills, rather than fluency and reading comprehension, which are captured in the broad reading scores.[30]

Overall, the results in Table 3 suggest that the *Cell-Ed* program significantly increased participants' Spanish reading outcomes over the four-month period, with an increase that is comparable to a child's increase in reading levels over 1.4 to 2.1 years. This does not imply that the impact of Cell-Ed is *equivalent* to 1.4 years of *in-school education* in Spanish, as the adult education program focused on narrower set of *reading* skills. With regards to reading skills per se, the broad reading measure is fairly broad, capturing skills beyond word recognition (e.g. fluency and reading comprehension) while excluding other advanced skills such as composition.

Comparing our effects with those of other rigorous evaluations of adult education outcomes is difficult, as such evaluations are very rare and often use different measures of learning. A large-scale evaluation of an adult education program in India finds that the programs increased learning outcomes by .06 – .15 s.d. (Banerji et al 2013). Aker et al (2012) find that a mobile phone-enhanced adult education program increased writing and math test scores by .20-.25 s.d. as compared with a traditional adult education program. The latter effect results from comparison between a traditional literacy program and the enhanced mobile phone-

[30]Table A4 shows results from median regression and show median impacts on basic and broad reading skills of 2.4 and 2.2 years respectively.

based program (and the effect size is thus lower than that of the mobile phone-enhanced literacy program itself). The simple (unreported) before-after comparison of the traditional literacy program in Aker et al. suggested a much larger effect size of over 5 s.d. of the baseline control group literacy scores. With this effect size, a sample of 70 observations was largely sufficient to determine whether *Cell-Ed* could be used to teach literacy skills without teachers.[31]

It is important to note that the sample in Banerji et al. (2015), Aker et al. (2012) and this current research are all drawn from differently truncated distributions of baseline outcomes, as the selection criteria for participation differed.[32] As a result, the normalization of effects into effect sizes will depend heavily on the standard deviation of outcomes in the control group and a comparison of effect sizes is not directly meaningful.[33] This is why our preferred outcome is the age equivalent measure, which allows us to contextualize the observed learning relative to a comparison group (children) that is more intuitive. The before-after comparison in Aker et al. regarding the traditional literacy and mobile phone-enhanced literacy program suggests that 8

[31]Specifically, the before and after values in the literacy program were .05 and 2.32, respectively, with standard deviations of .27 and 2.17. Using these values, the required sample size to achieve a power of 0.9 with a two-sided significance level of 0.05 is 20 observations in total (10 in each of treatment and control groups), even without including any other controls. The present study included a measurement of IQ and a much finer measure of literacy, so that the power calculations were quite conservative.

[32]The before-after comparison using data from Aker et al. (2012) relies on a sample population that targeted the most illiterate (and therefore had the most homogenous group of respondents and lowest standard deviation). The current sample included some respondents who were a bit more literate. The sample in Banerji et al. (2013) included some women who were able to read paragraphs (and thus had the largest variation in the control group at baseline).

[33] This also means that while reducing the baseline variation is an optimal strategy to maximize the power of the impact evaluation, it reduces the usefulness of comparisons across projects and programs.

months of literacy classes which took place three hours a day for 5 days a week led to an increase in literacy skills that was approximately equivalent to 2.1 and 2.3 years of in-school education in Niger. The impacts of *Cell-Ed* are of a similar magnitude (around 3.5 years for basic reading scores and between 1.4 and 2.1 years for broad reading scores), though they were achieved in a much shorter period of time, and with a much lower time investment on the part of the students. We believe the uniformity and high quality of the standardized *Cell-Ed* program is a main reason for the difference.

B. Dealing with Differential Attrition

Implementing the monotone treatment response assumption

The previous results ignore the differential attrition between the treatment and control groups, and hence provide biased estimates of the impact of the *Cell-Ed* program. We address attrition by providing a lower bound on the treatment effect by imputing the baseline value of test scores for non-respondents in the follow-up data (effectively assuming no treatment effect).[34] The resulting estimates are provided in Table 4A.[35]

On average, students assigned to the *Cell-Ed* treatment increased their basic reading scores by 3.6 age equivalent years, based on the simple difference in means, the difference in difference and the fixed effects estimates (Table 4A, Columns 1, 4 and 5). Controlling for baseline covariates (Column 2) and the baseline outcome (Column 3) increases the coefficient slightly, and the effect remains statistically significant. Turning to broad reading scores, the impact of *Cell-Ed* program ranges from 1.4 years (Column 6) to 2.1 years (Columns 7-10). The effects on both basic and broad reading scores are statistically significant at the 5 percent and 1 percent levels.[36]

Lee Bounds

---

[34]Since progressing on the platform is linked to skills acquisition and completion of quizzes, in theory, it would be possible to predict the test outcomes for non-respondents based on their progress on the platform. However, this would require that the online tests were completed by the intended respondent, which cannot be verified. It would also require a common support on platform progress between respondents and non-respondents, which is not the case in our context.

[35]The 95 percent confidence interval around these point estimates would correspond to non-parametric bounds. Because we also imputed baseline values for the control group non-respondents, these would not be true non-parametric bounds using the monotone treatment response assumption. Therefore, we prefer to provide the point estimates.

[36]A number of additional robustness checks to deal with non-random attrition were also conducted, including standard Lee bounds and adjusted Lee bounds. The results are largely robust to these specifications. The results are available upon request.

Table 4B shows both the traditional and adjusted Lee bounds for basic and broad reading scores. The relevant bounds on the treatment effects are in bold, namely, the lower bound of the 95% confidence interval (around the lower bound), as well as the upper bound of the 95% confidence interval (around the upper bound).[37] In all cases, the treatment effects are bounded away from zero: The point estimates of the lower bound for basic and broad reading scores are 4.43 and 2.08 years, respectively, with a statistically significant effect at the 1 percent level (Panel A, Columns 1 and 4). Focusing on the conservative lower bound of the 95% confidence interval, the treatment effect is bounded below by 1.58 and 0.72 years for basic and broad reading scores, respectively (Panel A, Columns 2 and 5).

C. Impacts of the Cell-Ed Program on Empowerment

The previous results have shown that the *Cell-Ed* program increased students' basic and broad reading scores in a relatively short period of time. Yet beyond learning, a key claim of many adult education programs is one of empowerment: that educated adults are able to make better decisions, lead more independent lives and search for jobs that require literacy skills. This section investigates whether the increased learning due to *Cell-Ed* led to improvements in self-esteem and self-efficacy, using the RSES and the GSES.

[37]The lower bound of the 95% confidence interval of the lower bound and the upper bound of the 95% confidence interval of the upper bound are conservative estimates of the 95% confidence interval around the estimate (Lee, 2009). As noted above, we bootstrap the confidence intervals.

Table 5 reports the results of the same specifications as those in Table 3, using measures of self-esteem and self-efficacy as the dependent variables. Overall, the *Cell-Ed* program increased students' self-esteem score by 2.49 points, with a statistically significant effect at the 5 percent level (Table 5, Column 1). These results are robust to the inclusion of additional covariates (Column 2), as well as a value-added specification (Column 3). Yet the results are no longer statistically significant once the DD (Column 4) and fixed effects specifications (Column 5) are used. Table A5 documents similar results using the weekly calls; in these regressions, the estimated impacts are statistically significant at the 5 percent level in all specifications.

Turning to self-efficacy, there is no statistically significant impact of the *Cell-Ed* program on self-efficacy for the first three specifications (Columns 6-8). This is perhaps unsurprising, as baseline self-efficacy scores were relatively higher in the control group. Using DD and fixed effects specifications (Columns 9 and 10), the *Cell-Ed* program is associated with an increase of 3.75-4.75 points in students' self-efficacy scores, with a statistically significant effect at the 10 and 1 percent levels, respectively.[38,39] The non-parametric bounds in appendix Table A7 show that we cannot bound the treatment effect for empowerment away from zero. Overall, these

[38]Table A6 shows that these results are robust to including baseline mobile ownership and self-esteem, as well as using normalized empowerment scores as the dependent variable.

[39]Table A7 uses an empowerment index comprising the normalized RSES and the normalized GSES as the dependent variable. The results are broadly consistent with those in Table 6.

results suggest that the *Cell-Ed* program may be associated with an increase in participants' self-esteem, but this is not a robust conclusion. We are also unable to conclude that there is a statistically significant effect on self-efficacy.

We are also interested in the relationship between the *Cell-Ed* program, learning and self-esteem over time. Table A7 investigates this relationship using data from the weekly self-esteem measurements and the *Cell-Ed* platform. Overall, self-esteem is negatively associated with lack of learning progress during the previous week (Column 1). In other words, as a student's proportion of incorrect responses during the previous week increases, the student's self-esteem decreases. When disaggregating these results by gender (Columns 2-4), the relationship is driven entirely by men. This suggests that while overall self-esteem is higher for *Cell-Ed* participants by the end of the program, perceptions of self-esteem may change over time, particularly when experiencing learning failures.

While the results in Table A7 provide important insights into the dynamic relationship between learning and self-esteem, Table A9 shows how learning is associated with the likelihood of dropping out of the program. *Cell-Ed* participants who had a higher proportion of errors over the course of their interactions with the platform were more likely to stop interacting with the platform and drop out of the subsequent survey round (Column 1). Columns 2 and 3 focus on the sub-sample of women: female participants with a proportion of errors exceeding a threshold of 40% were more likely to drop out of the program. The effects seem to be particularly strong among participants with a high proportion of errors, as suggested by the quadratic specification (Columns 4-6).[40]

V. Mechanisms

Why might the *Cell-Ed* program improve students' learning outcomes? A key hypothesis of the program was that a mobile phone-based program might reduce the opportunity costs of investing in adult education for busy adults by allowing them to learn when, where and how they wished. We are able to test this hypothesis by using the *Cell-Ed* platform data, which provides information on when *Cell-Ed* students interacted with the platform.

Figures A4, A5, and A6 show the learning patterns of *Cell-Ed* participants, by employment status. Overall, participants learned at all times of the day and night, at times much later than a standard adult education class would operate, suggesting that the mobile-based course was more appropriate for adults' work schedules (Figure A3). *Cell-Ed* participants spent less time on the platform during weekends as compared with weekdays (Figure A4). Finally, Figure A5 shows the distribution of interactions with the platform in terms of the amount of time spent: most interactions are quite short, approximately 10 minutes, suggesting

[40]We note that the sample for men is too small for robust results, as only 12.5 percent of men drop out (as compared to 1/3 of women). Table A9 also shows that there is no evidence that dropping out of the program is related to lower self-esteem, at least not in the weeks leading up to the dropout. Non-respondents have slightly higher levels of self-esteem at baseline and there does not seem to be any change in their self-esteem over the course of the program, in particular not in the last three weeks of the program.

that learners use the platform for short learning episodes. This is in contrast to most adult education programs, which hold classes for several hours on a weekly basis. On average, though, time spent on learning is similar to standard literacy programs: *Cell-Ed* learners were spending an average of 2.95 hours per week (3.75 for respondents) on the platform. The target for in-situ learning is about 4-6 hours per week. Accounting for in-situ absences, the 50 percent dropout rate and travel time suggests that *Cell-Ed* learners spent slightly more time (on average) on literacy acquisition than students in a standard adult education program. While we do not know whether this flexibility is indeed the key to the large improvements in reading outcomes, the usage patterns (and different usage patterns by employment status) highlight that people did make use of it.

VI. Conclusion

Information technology, and in particular mobile phones, enables individuals to access information at any time and location. Yet despite the potential, there is limited evidence that simple mobile phone devices can be used to teach basic educational skills, such as literacy and numeracy. Using a randomized controlled trial of a mobile phone-based adult education program (*Cell-Ed*), we find that the program significantly increased adult students' reading levels within a short period of time. These results are robust to accounting for significant non-random attrition using non-parametric methods. We also find that the *Cell-Ed* program may be correlated with an increase in students' self-esteem by the end of the program, and that there is a dynamic relationship between learning progress and self-esteem.

Admittedly, our experimental set-up has several limitations. First, we are unable to compare learning via the *Cell-Ed* platform with learning in a traditional adult education program, as – to our knowledge – there is no benchmark data for learning in traditional adult literacy programs in the US. A key concern to policy-makers should be to provide such information in an easily accessible manner. We were also not able to provide evidence on an interaction between the *Cell-Ed* platform and traditional literacy programs. As a result, we are unable to conclude whether such programs are complements or substitutes for teachers and in-classroom learning (Linden, 2008). Second, our small sample size limits the external validity of our results, though these worries are somewhat mitigated by the fact that the impacts also hold (in a before-after comparison) when the control group is phased in. Overall, our results show that a distance learning program via a simple mobile phone significantly improved adults' learning outcomes in this context, and suggests that there is great scale and scope for using these technologies in education programs in both high and low-income countries.

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.

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, N., Bergman, P. & Matsheng, M. "Experimental evidence on learning using low-tech when school is out." *Nat Hum Behav* **6**, 941–950 (2022).

Ashraf, Nava, Dean Karlan, and Wesley Yin. March 2010. "Female Empowerment: Further Evidence from a Commitment Savings Product in the Philippines." *World Development* 38, no. 3 (March 2010): 333–344.

Ashraf, Nava. September 2009. "Spousal Control and Intra-Household Decision-Making: An Experimental Study in the Philippines." *American Economic Review* 99, no. 4 (September 2009): 1245–1277.

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, R., Berry, J. and Shotland, M., 2017. "The impact of mother literacy and participation programs on child learning: evidence from a randomized evaluation in India." *AEJ: Applied Economics.* 9(4).

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.

Carlana, M and Eliana la Ferrara. 2021. "Apart but Connected: Online Tutoring and Student Outcomes during the COVID-19 Pandemic" (with M. Carlana), CEPR Discussion Paper 15761, 2021.

Carron, G. 1990. "The Functioning and Effects of the Kenya Literacy Program." *African Studies Review*, pp. 97-120.

Case, Anne. 2005. "The Primacy of Education," In *Understanding Poverty,* ed. Abhijit Vinayak Banerjee, Roland Benabou and Dilip Mookherjee. Oxford ; New York: Oxford University Press.

Banerji, R., Berry, J and Shotland, M., 2017. "The impact of mother literacy and participation programs on child learning: evidence from a randomized evaluation in India." *AEJ: Applied Economics.* 9(4).

Duflo, Esther. 2012. "Women Empowerment and Economic Development." *Journal of Economic Literature*. 50(4). 1051-1079.

Doepke, Mathias and Michele Tertilt. 2014. "Does Female Empowerment Promote Economic Development?" *NBER Working Paper 19888,* NBER, Inc.

Kim, Y., R. Telang, W. Vogt and R. Krishnan. 2009. "An Empirical Analysis of Mobile Voice Service and SMS: A Structural Model." *Management Science*, 56(2), pp. 234-52.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects" *The Review of Economic Studies*, 6, 1072-1102.

Linden, Leigh. 2008. "Complement or substitute? The effect of technology on student achievement in India." JPAL Working Paper.

Literacy Network of Greater LA and United Way. 2004. Literacy @ work. Skills Today, Jobs Tomorrow. Los Angeles: Literacy Network of Greater LA and United Way.

Manski, Charles. 1997. “Monotone Treatment Response,” *Econometrica*, 65, 1311–1334.

McGrew, Kevin, Frederick Schrank and Richard Woodcock. 2007. *Techinical Manual. Woodcock-Johnson III Normative Update.* Rolling Meadows. IL: Riverside Publishing.

National Center for Education Statistics. 2003. "The 2003 National Assessment of Adult Literacy", Washington DC, USA: National Centre for Education Statistics, U.S. Department of Education.

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.*

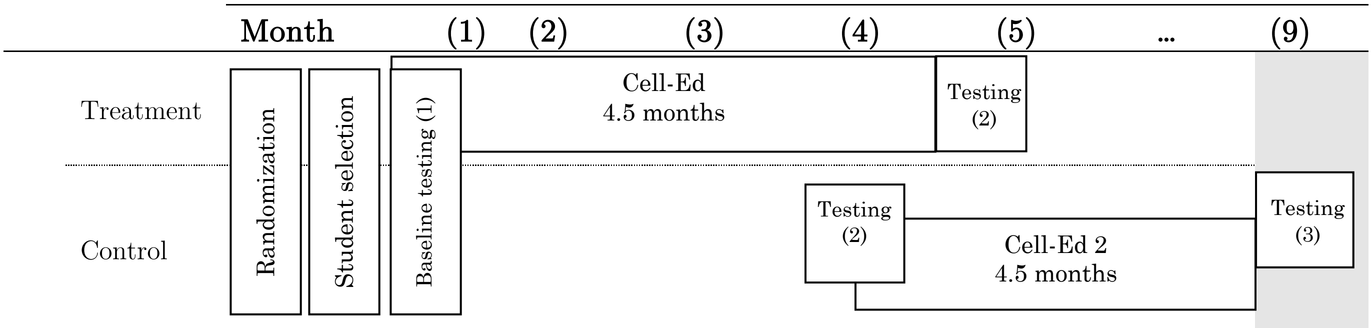
Oxenham, John, Abdoul Hamid Diallo, Anne Ruhweza Katahoire, Anne Petkova-Mwangi and Oumar Sall. 2002. *Skills and Literacy Training for Better Livelihoods: A Review of Approaches and Experiences*. Washington D.C.: World Bank.

Romain, R. and L. Armstrong. 1987. *Review of World Bank Operations in Nonformal Education and Training*. World Bank, Education and Training Dept., Policy Division.

UNESCO. 2005. *Education for All: Global Monitoring Report. Literacy for Life*. Paris: UNESCO.

UNESCO. 2012. *Education for All: Global Monitoring Report. Youth and Skills: Putting Education to Work*. Paris: UNESCO.

Figure 1. Timeline of Data Collection and Adult Education Activities



Notes: This figure represents the timeline for the *Cell-Ed* program. Testing (1), Testing (2) and Testing (3) refer to the first, second and third round of testing, respectively.

Figure 2. Lee Bounds and attrition

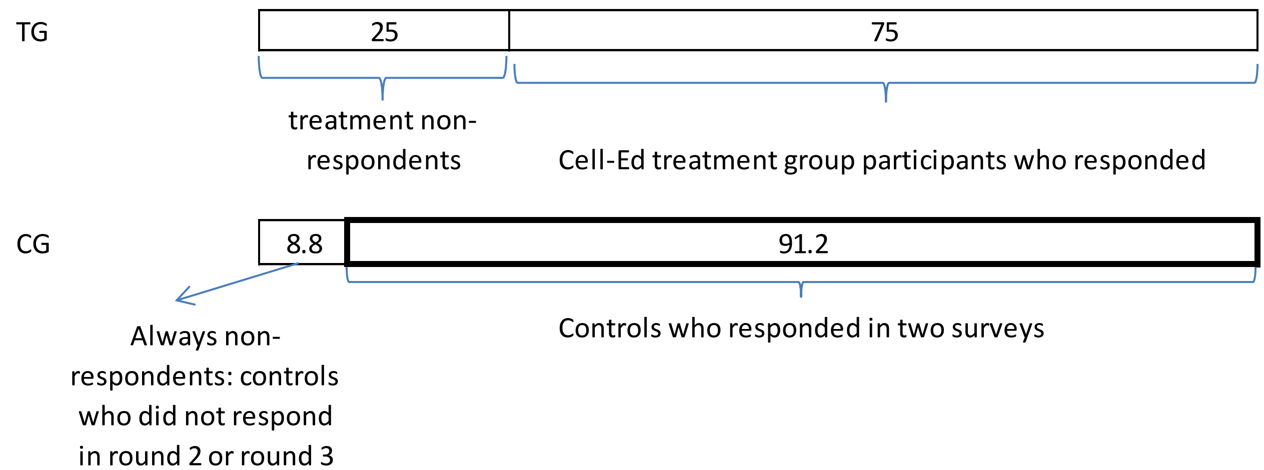
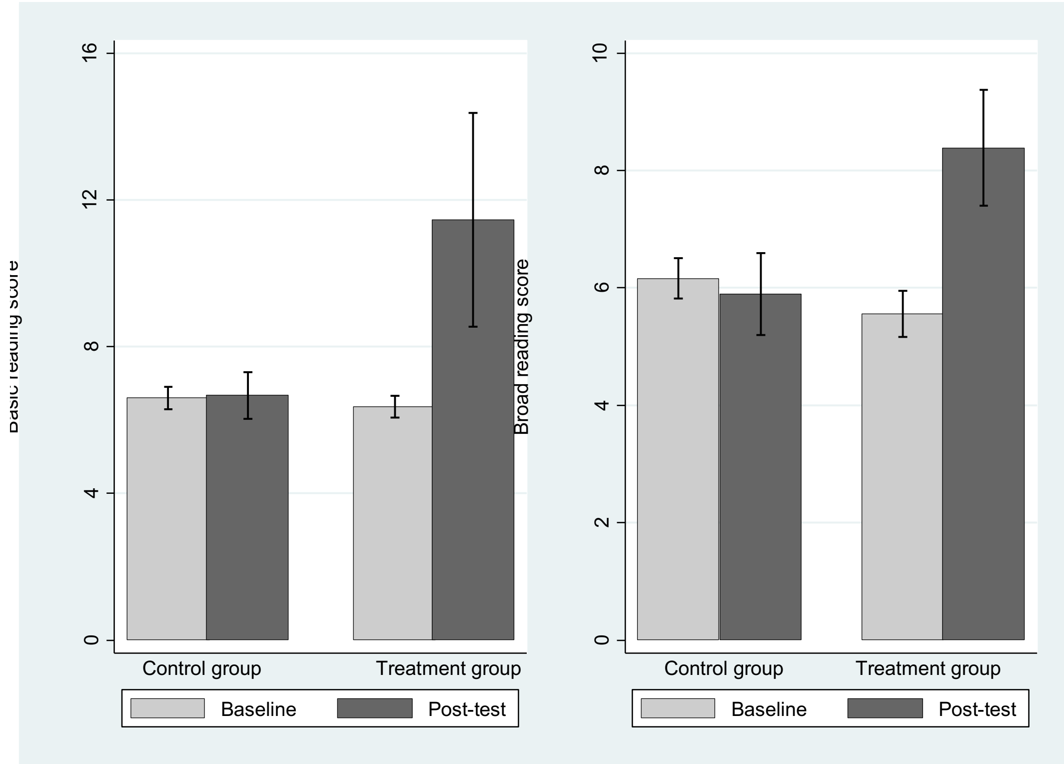


Figure 3: Basic and Broad Reading Scores (by Treatment Status and Survey Round)



Notes: Graph pictures the means scores on the Woodcock-Munoz III reading assessment, for basic (left panel) and broad (right panel) reading scores, by treatment status. Baseline survey measurement means are in lighter color, Round 2 measurements in darker color. Also depicted are the 95% confidence intervals around the mean.

Table 1: Baseline Balance

	Cell-Ed	Initial Control	
	Mean (s.d.)	Mean (s.d.)	Difference Coeff (s.e.)
	(1)	(2)	(3)
Panel A: Covariates			
Age in years	48.50	46.70	1.81
	(12.50)	(13.13)	(3.06)
Male	0.28	0.21	0.07
	(0.45)	(0.41)	(0.10)
Verbal IQ test	7.74	8.02	-0.28
	(1.85)	(2.31)	(0.50)
Have you ever attended formal school?	0.47	0.56	-0.09
	(0.51)	(0.50)	(0.12)
For how many years did you attend school?	0.80	1.66	-0.85
	(1.10)	(3.18)	(0.58)
Are you currently employed?	0.39	0.35	0.04
	(0.49)	(0.49)	(0.12)
Do you currently own a cell phone?	0.61	0.41	0.20*
	(0.49)	(0.50)	(0.12)
In a normal day, do you make cell phone calls?	0.64	0.67	-0.03
	(0.49)	(0.48)	(0.12)
Panel B: Baseline Outcomes			
Basic reading test, age equivalent	6.36	6.60	-0.24
	(1.27)	(1.55)	(0.34)
Broad reading test, age equivalent	5.56	6.16	-0.60
	(1.66)	(1.79)	(0.42)
Rosenberg Self-Esteem Scale	19.52	18.03	1.49*
	(3.61)	(2.52)	(0.76)
General Self-Efficacy Score	32.88	35.03	-2.15
	(5.97)	(6.05)	(1.47)
Number of observations	36	34	70

Notes: Column 1 presents the mean (and standard deviation) of baseline characteristics for the Cell-Ed treatment group, Column 2 presents the mean (and standard deviation) of baseline characteristics for the control group. Column 3 reports the difference between the two. ***, **, * denote statistical significance at the 1, 5, 10 percent levels, respectively.

Table 2A: Survey Non-response

	Cell-Ed		Control		
	(1)	(2)	(3)	(4)	(5)
Round 1: Baseline	36	100%	34	100%	
Round 2: Endline for treatment group	27	75%	31	91%	100%
Round 2: Endline for control/phase-in treated	NA		21	62%	68%

Notes: The rows represent the three survey rounds. Column 1 presents the number of observations for the Cell-Ed treatment group, Column 2 the percentage of treatment observations in that survey round relative to the baseline number (36). Column 3 presents the number of observations for the control group (and later phase-in group). Columns 4 contains the percent control group members observed relative to baseline (34) Column 5 contains the percent observations surveyed relative to the Round 2 (which can be thought of as a second baseline for the phase-in control group), namely 31.

Table 2B: Interaction with Platform

	Respondents				Non-Respondents			
	mean	sd	min	max	mean	sd	min	max
Panel A: Treatment Group								
Days between first contact and last contact with platform	123.78	29.39	48	150	92.78	45.65	41	150
Time spent on the platform (minutes)	3144.44	2194.56	469	8886	466.56	515.92	100	1695
Modules completed	401.56	74.48	184	436	31.11	29.16	4	104
Number of observations	27				9			
Panel B: Control/phase-in treatment								
Days between first contact and last contact with platform	133.71	25.47	54	150	103.90	52.71	20	150
Time spent on the platform (minutes)	2041.14	1183.73	432	3780	707.30	851.19	62	2594
Modules completed	328.33	123.75	104	436	81.40	93.29	2	265
Number of observations	21				10			

Notes: Table presents statistics related to interactions of learners with the Cell-Ed platform. Panel A has outcomes for phase 1 when the treatment group was treated. Panel B has outcomes for phase 2 when the control group was treated (this excludes the 3 observations who were never put on the platform). The four left columns provide information (mean, standard deviation, minimum, maximum) for respondents and the four right columns for non-respondents. Days between first and last contact with the platform is the number of days between the first and the last interaction (+1), which is censured for research purposes above at 150 (in practice, participants could continue on the platform after the respective endline survey). Time spent on the platform is the total hours spent on the Cell-Ed platform. Modules completed are the total number of modules completed out of 436.

Table 3: OLS Regression Results

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Basic Reading					Broad Reading				
	simple means	incl covariates	value-added	DD	FE	simple means	incl covariates	value-added	DD	FE
Treatment * Post				4.960***	4.763***				3.003***	2.767***
				(1.434)	(1.383)				(0.681)	(0.452)
Treatment group	5.653***	5.539***	5.840***	-0.024		2.582***	2.533***	3.096***	-0.757**	
	(1.854)	(1.789)	(1.750)	(0.477)		(0.679)	(0.641)	(0.556)	(0.373)	
Post				0.150	0.126				-0.212	-0.177
				(0.465)	(0.145)				(0.442)	(0.132)
Y (t-1)			1.146***					0.909***		
			(0.303)					(0.122)		
Age in years		0.003	-0.019	0.006			0.009	-0.016	0.010	
		(0.047)	(0.042)	(0.024)			(0.024)	(0.013)	(0.014)	
Male		0.524	0.112	0.465			0.243	-0.324	0.367	
		(1.862)	(1.744)	(0.932)			(0.847)	(0.558)	(0.489)	
Verbal IQ test		0.737*	0.562	0.428**			0.326*	0.112	0.283***	
		(0.387)	(0.348)	(0.186)			(0.174)	(0.109)	(0.099)	
Observations	58	58	58	128	128	58	58	57	127	127
R-squared	0.275	0.345	0.415	0.321	0.322	0.315	0.384	0.706	0.363	0.558

Notes: Results from a regression of observed reading test outcomes on different sets of covariates for basic reading scores (Columns 1-5) and broad reading scores (Columns 6-10). Columns 1 and 6 provide estimates from a regression just on the treatment dummy and strata fixed effects. Columns 2 and 7 include covariates (age, gender and IQ). Columns 3 and 8 include the baseline outcome variable denoted by Y(t-1), which is baseline basic reading score for Column 3 and the baseline broad reading score for Column 8. Columns 4 and 9 provide difference-in-difference specifications. The number of observations for the DD specifications (Columns 4 and 9) is higher, as the baseline scores of later control and treatment non-respondents are included. In the value-added and DD specifications for the broad reading score, one observation is missing as the interviewer erroneously stopped the tests after the components of the basic reading score had been completed. Columns 5 and 10 present results from a fixed effects specification. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1, 5, 10 percent levels, respectively.

Table 4A: OLS Regression Results (With Baseline Imputation for Missing Observations)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Basic Reading					Broad Reading				
	simple means	incl covariates	value-added	DD	FE	simple means	incl covariates	value-added	DD	FE
Treatment * Post				3.551***					2.162***	
				(1.196)					(0.642)	
Treatment group	3.574**	3.759**	4.408***	-0.141	3.551***	1.375**	1.455**	2.239***	-0.824**	2.086***
	(1.475)	(1.510)	(1.455)	(0.453)	(1.099)	(0.624)	(0.617)	(0.506)	(0.369)	(0.394)
Post				0.115	0.115				-0.162	-0.162
				(0.447)	(0.132)				(0.426)	(0.121)
Y (t-1)			1.324***					0.961***		
			(0.292)					-0.132		
Age in years		-0.022	-0.040	-0.004			-0.007	-0.025*	0.003	
		(0.044)	(0.041)	(0.023)			(0.021)	(0.014)	(0.013)	
Male		1.478	0.800	0.995			0.860	0.140	0.675	
		(1.694)	(1.560)	(0.903)			(0.782)	(0.513)	(0.474)	
Verbal IQ test		0.692*	0.491	0.422**			0.320*	0.117	0.281***	
		(0.378)	(0.335)	(0.191)			(0.172)	(0.119)	(0.100)	
Observations	70	70	70	140	140	70	70	69	139	139
R-squared	0.144	0.238	0.334	0.231	0.242	0.143	0.249	0.605	0.277	0.415

Notes: Results from a regression of observed or imputed reading test outcomes on different sets of covariates for basic reading scores (Columns 1-5) and broad reading scores (Columns 6-10). For observations with missing post-treatment outcomes, baseline values are imputed. Columns 1 and 6 provide estimates from a regression just on the treatment dummy and strata fixed effects. Columns 2 and 7 include covariates (age, gender, and IQ). Columns 3 and 8 include the baseline outcome variable denoted by Y(t-1), which is baseline basic reading score for Column 3 and the baseline broad reading score for Column 8. Columns 4 and 9 provide difference-in-difference specifications. Note that the number of observations for the DD specifications (Columns 4 and 9) is higher, as the baseline scores of later control and treatment non-respondents are included. In the value-added and DD specifications for the broad reading score, one observation is missing as the interviewer erroneously stopped the tests after the components of the basic reading score had been completed. Columns 5 and 10 present results from a fixed effects specification. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1, 5, 10 percent levels, respectively.

Table 4B: Lee Bounds and Adjusted Lee Bounds for the Treatment Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Basic Reading			Broad Reading		
	Point estimate	Confidence interval lower bound	Confidence interval upper bound	Point estimate	Confidence interval lower bound	Confidence interval upper bound
Lower bound	4.433***	**1.583**	7.280	2.032***	**0.723**	3.143
Upper bound	5.437***	2.604	**8.250**	3.169***	1.824	**4.408**
Observations						
Total	70			70		
Attrition	12			12		
Trimmed	6			6		

Notes: Table presents non-parametric bounds, based on Lee (2009) for basic reading and broad reading. We present estimates for the lower and the upper bound. Columns 1 and 4 contain the point estimate for basic and broad reading scores, respectively. Columns 2 and 5 contain the lower bound of the 95% confidence interval and columns 3 and 6 the upper bound of the 95% confidence interval for basic and broad reading scores, respectively. All confidence intervals are bootstrapped. The lower 95% CI bound and the upper 95% CI bound are overly conservative bounds on the treatment effect (as Imbens and Manski (2004) show), so are wider than the 95% confidence interval for the treatment effect. ***, **, * denote statistical significance at the 1, 5, 10 percent levels, respectively.

Table 5: OLS Results for Empowerment

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Rosenberg Self-Esteem Score					General Self-Efficacy Scale				
	simple means	incl covariates	value-added	DD	FE	simple means	incl covariates	value-added	DD	FE
Treatment * Post				1.115					3.747*	
				(1.270)					(2.203)	
Treatment group	2.490**	2.572**	2.601**	1.115	1.379	1.422	1.325	3.132	-2.264	4.757***
	(1.099)	(1.053)	(1.235)	(0.766)	(1.395)	(2.103)	(1.912)	(2.263)	(1.474)	(1.693)
Post				1.748**	1.621**				-0.656	-0.966
				(0.792)	(0.757)				(1.664)	(1.357)
Y (t-1)			-0.060					0.463**		
			(0.268)					(0.194)		
Age in years		0.040	0.038	0.002			0.092	0.051	0.090*	
		(0.045)	(0.049)	(0.025)			(0.065)	(0.071)	(0.049)	
Male		0.810	0.893	1.333*			0.442	0.173	0.001	
		(1.246)	(1.509)	(0.735)			(2.061)	(2.043)	(1.318)	
Verbal IQ test		0.494	0.512	0.335*			-0.053	-0.213	0.055	
		(0.351)	(0.384)	(0.186)			(0.616)	(0.564)	(0.343)	
Observations	55	55	53	122	122	55	55	53	122	122
R-squared	0.224	0.285	0.284	0.294	0.192	0.068	0.103	0.296	0.099	0.152

Notes: Results from a regression of observed reading test outcomes on different sets of covariates for the Rosenberg Self-esteem score (Columns 1-5) and the General self-efficacy Score (Columns 6-10). Columns 1 and 6 provide estimates from a regression just on the treatment dummy and strata fixed effects. Columns 2 and 7 include covariates (age, gender and IQ). Columns 3 and 8 include the baseline outcome variable denoted by Y(t-1), which is baseline self-esteem score for Column 3 and the self-efficacy score for Column 8. Columns 4 and 9 provide difference-in-difference specifications. Note that the number of observations for the DD specifications (Columns 4 and 9) is higher, as the baseline scores of later control and treatment non-respondents are included. Columns 5 and 10 present results from a fixed effects specification. Relative to Table 3, 6 observations are missing for the DD specifications, as some respondents could not understand and could not complete the measures for self-esteem and self-efficacy. This also reduces the mean difference specifications and the value-added specifications. Robust standard errors in parentheses. ***, **, * denote statistical significance at the 1, 5, 10 percent levels, respectively.