The Impact of a Computer-Based Adult Literacy Program on Literacy and Numeracy: Evidence from India

Ashwini Deshpande
Department of Economics, *Delhi School of Economics, University of Delhi, Delhi 110007, India*
Email: ashwini@econdse.org

Alain Desrochers
School of Psychology, University of Ottawa,
136 Jean-Jacques-Lussier, Ottawa, Ontario, Canada K1N 6N5
Email: damch@uottawa.ca

Christopher Ksoll
Mathematica Policy Research
505 14th Street, Oakland, CA 94612, USA
Email: cksoll@mathematica-mpr.com

Abu S. Shonchoy (Corresponding Author)
Institute of Developing Economies-JETRO and New York University
Wagner School of Public Services, 295 Lafayette Street, NY 10025, USA
Email: parves.shonchoy@gmail.com, Phone: +1 (917)972-0041, Fax: +1 (212)995-4162

The Impact of a Computer-Based Adult Literacy Program on Literacy and Numeracy: Evidence from India

Abstract

With over 700 million illiterate adults worldwide, governments in many developing countries have implemented adult literacy programs. Typically these program have low rates of success partly because the quality of teaching is heterogeneous. Standardization of teaching provided by computer-aided instruction might be a solution. However, there is little rigorous evidence of the effectiveness of computer-based adult literacy programs in delivering high-quality literacy and numeracy in the developing world. To fill this void in the literature, we study the impact of a computer-based adult literacy program, Tara Akshar Plus, on the literacy and numeracy skills of previously illiterate adult women in the north Indian state of Uttar Pradesh. Through a randomized control trial, we measure learning outcomes with individual-level literacy and numeracy tests and find statistically significant positive impacts of this computer-aided program on literacy and numeracy outcomes of women who undergo the TARA Akshar Plus program – relative to the control group. The effects are statistically significant but small in magnitude for women who were entirely illiterate prior to the program. The learning impacts are substantially larger for learners who knew at least a handful of letters at the beginning of the program. We compare the improvement in learning to that of another adult literacy and numeracy program. We conclude that TARA Akshar Plus is the more effective of the two, but the literacy and numeracy level achieved are not large enough to make many entirely illiterate learners become functionally literate.

JEL: I20, J16, O53
Keywords: Adult Literacy Program; Adult Education; ICT; Women; India

1. Introduction

The overwhelming majority of the world's 757 million illiterate adults live in developing countries (UIS, UNESCO, 2015). Adult literacy programs aim to improve the skills, and hence the earning potential and other socioeconomic outcomes, of illiterate adults. However, traditional adult literacy programs, typically operated by governments, have been largely ineffective due to low enrollment, high dropout rates, and rapid skill depreciation (Abadzi, 1994, 2003; Oxenham, 2002; and Ortega & Rodríguez, 2008). Recent advances in adult literacy programs have sought to integrate modern information and communication technology (ICT) into effective teaching methods (for an insightful overview, see, Wagner & Kozma, 2005). The use of modern ICT, which comprises computers, mobile phones, and tablets, could improve the quality and effectiveness of learning with the aid of interactive tools, the use of animation, and the implementation of effective teaching principles (Iftekar & Hyeon, 2016).

There is little rigorous evidence of the effectiveness of computer-based adult literacy programs in delivering high-quality literacy and numeracy teaching in the developing world. (Berger (2001) provides a review of the effectiveness computers in adult literacy classes in the US.) By rigorous we mean evidence that allows for the causal attribution of the observed effect to the literacy program, which at a minimum requires a deliberate attempt to account for confounding factors, for example by creating a control or comparison group.

In this paper, we seek to fill this void in the literature by investigating the impact of TARA Akshar Plus (TA+), a systematic computer-based adult literacy and numeracy program conducted in the state of Uttar Pradesh in northern India, on the literacy and numeracy of neo-literate adult women. TA+ has been implemented by the Indian NGO Development Alternatives (DA),[1] which claims that the success rate of this program is over 90 percent.[2] Despite having reached more than 100,000 participant learners, the program has never been scientifically evaluated. Our study provides the first thorough assessment of TA+ and one of the first rigorous studies of the effectiveness of a computer-based adult literacy program on literacy and numeracy within the developing world.

We employ the random assignment of 717 illiterate women from 18 villages into two groups: (i) a treatment group that could undergo the TA+ program immediately and (ii) a control group that could undergo the program only at a later date. Respondents in the treatment and control groups were tested pre- and post-intervention. We present results on the impact of TA+ on learning outcomes by combining the random assignment with individual-level test results.

Our main results are as follows. We find that TA+ has a statistically significant effect on literacy and numeracy in the short run, especially in basic literacy and numeracy skills such as reading letters and words and counting and number recognition. However, the effects on more advanced dimensions of reading and numeracy, such as reading paragraphs and addition/ subtraction, are small and suggest that not many learners become functionally literate.

The rest of the paper is organized as follows: section 2 contains a brief review of related literature; section 3 gives the background to the TA+ program; section 4 describes the design of the experiment; section 5 describes the data and presents descriptive statistics; section 6 presents results of the impact analysis and compares the learning effects of TA+ to results from similar programs; and section 7 concludes the paper.

2. A Brief Review of Related Literature

Impact assessments of adult literacy programs can be grouped into two kinds: one set that measures the direct effects, namely, the acquisition of literacy or numeracy, and the other that measures the indirect or extended effects, such as intrahousehold sharing or child health outcomes. The set that measures the direct or immediate impacts of adult literacy programs, namely, acquisition of literacy and/or numeracy, primarily consists of studies that suffer from some or all of the following problems: very small sample sizes, flawed experimental design (e.g., lack of a comparison group), and poorly designed assessment tools (Carron, 1990; Ortega & Rodríguez, 2008). An exception is Banerji, Berry, & Shotland (2015) who provide a rigorous evaluation of literacy classes on language and math scores in the states of Bihar and Rajasthan in India.

The studies that measure the indirect or extended effects of adult literacy programs include the assessment of a large-scale Ghanaian adult literacy program on household consumption (Blunch & Pörtner, 2011); the assessment of maternal participation in adult literacy programs on child mortality in rural Ghana (Blunch, 2013); and the assessment of the positive impact of maternal literacy on children's math scores in India (Banerji et al., 2015). These studies reveal that adult literacy programs can have a positive effect on participants' literacy, though the increases are small in magnitude. Adult literacy programs have also been evaluated (mainly through a comparison of literate and illiterate adults) in terms of their impact on outcomes such as interhousehold sharing (Maddox, 2007), individual earnings (Basu et al., 2001), and children's height-for-age (Gibson, 2001). However, the results of these comparisons are not necessarily attributable to specific adult literacy programs.

The broader literature on education highlights the robust positive correlations between education and several desirable socioeconomic outcomes. The positive correlation between maternal education and child outcomes has been well documented (see, among others, Strauss & Thomas, 1995; Gakidou, Cowling, Lozano, & Murray, 2010; Paxson & Schady, 2007; White & Masset, 2003; Senauer, Garcia, & Jacinto, 1988; Senauer & Garcia, 1991; Haddad, 1999; Thomas, 1990; Hopkins, Levin, & Haddad, 1994). Higher education levels for women and girls are also positively correlated with lower fertility; improved health, hygiene, and education; better saving practices; and increased gender equity (Senauer et al., 1988; Thomas, 1990; Hopkins et al., 1994; Strauss & Thomas, 1995; Haddad, 1999; White & Masset, 2003; Paxson & Schady, 2007; Gakidou et al., 2010).

Turning to ICT-based or ICT-enhanced adult literacy programs in developing countries, there is generally little information about their impacts. A study conducted by Aker, Ksoll, and Lybbert (2012) provided evidence that teaching completely illiterate adults to use mobile phones within the context of adult literacy classes in rural Niger can increase math and reading test scores by 25 percent. Chugdar (2014) reported basic cross-tabulations from a survey of 409 illiterate adults in semi-urban locations in Gujarat, India, that highlighted the potential role of mobile phones in addressing adult literacy. Also in the Indian context, Wagner, Daswani and Karnati (2010) study the impact of a computer-assisted mother-tongue literacy program among out-of-school youth in the state of Andhra Pradesh. Using a quasi-experimental design, they find modest but promising impacts of this program on learning outcomes.

The literature on the effectiveness of computer-assisted learning in schools is large enough that a number of meta-analyses have been conducted, though these largely comprise studies in developed country settings. Tamim et al. (2011) provide a second-order meta-analysis – a meta-analysis of meta-analyses – that finds that the use of computer technology has an average effect size of 0.33-0.35 standard deviations on student learning. The effect sizes are, however, reduced to about half the size when only rigorous studies are included in the meta-analysis. Cheung and Slavin (2012) provide a meta-analysis that focusses on reading outcomes and find that technology has an average effect size of 0.16 standard deviations on reading outcomes. Using similarly restrictive inclusion criteria for studies, Cheung and Slavin (2013) find an average effect size of math-focused programs on math outcomes of 0.15 standard deviations.

The literature on the impacts of computer-assisted learning on student outcomes in developing countries is rapidly growing. It reveals mixed finding related to the effectiveness of computer-assisted learning. Barrera-Osorio and Linden (2009) and Beuermann *et al.* (2015), find either no or mixed effects of the introduction of computers in schools on children's learning outcomes (e.g., test scores) or cognitive skills in a range of settings. These limited results may be due to the failure to incorporate computers into the educational process. On the other hand, Mo *et al.* (2013) examine the "One Laptop per Child (OLPC)"[3] program in migrant schools in Beijing schools and find that six months of access to a computer improved standardized math scores by 0.17 standard deviations, as well as raising student scores on a computer skills scale by 0.33 standard deviations. Lai *et al.* (2013) study a remedial computer-assisted learning program in minority schools in rural China and find that it increases learning by 0.14-0.20 standard deviations. Banerjee *et al.* (2007)

find that a computer-assisted learning program involving math-focused games improved math scores among schoolchildren in India by 0.47 standard deviations.

3. Background

About one-third of the world's illiterate population lives in India, where illiteracy affects primarily girls and women. The female literacy rate—defined as including all females ages 7 and over who can read and write—is 65 percent, whereas the male literacy rate is over 80 percent. Females constitute 64 percent of the total illiterate population ages 7 and above (Census, 2011). Although India has been successful in raising the primary enrollment rates of boys and girls through programs such as *Sarva Shiksha Abhiyan* and the midday meal scheme (raising the net primary enrollment rate to 93 percent in 2011[4]), progress on adult literacy has been comparatively limited (Kapur & Murthi, 2011).

The National Literacy Mission (NLM) was launched in 1988 to increase functional literacy among 15- to 35-year-olds in India. Because these individuals are in the "productive and reproductive period of life," the NLM offers them "a second chance" at functional literacy. The Indian government conducts literacy programs for both their intrinsic and instrumental values. Thus, the NLM defines functional literacy as a composite set of goals: "self-reliance in the 3 R's [reading, writing, arithmetic]; becoming aware of the causes of deprivation and moving towards the amelioration of their condition by participating in the process of development; skill development to improve economic status and general well-being; and imbibing values of national integration, conservation of environment, women's equality, and observance of small family norms, etc." [5,6]

The NLM initially aimed to increase literacy rates to 75 percent by 2007. To date, however, only the state of Kerala has been declared "totally literate." This slower-than-anticipated progress has been attributed to large class sizes, inflexible schedules, poorly designed curricula, and, consequently, low participation. Partially in response to concerns regarding the poor quality of existing programs, the Indian government has extended its approach by supporting e-learning programs, including the TA+ literacy and numeracy program.

TA+ is one of the few e-learning programs accredited by the NLM. TA+ was developed by the UK-based company ReadingWise, which uses interactive computer-based learning modules. TA+ currently comprises two modules: TARA Akshar, a 36-day literacy program, and TARA Ganit, an 18-day numeracy program. TA+ began operating in 2005. As of 2015, more than 150,000 women had participated in its adult literacy classes, with a reported success rate of 95 percent. The TA+ program is funded by both the Indian government and international organizations. The learners we studied were part of an IKEA funded program in the eastern part of Uttar Pradesh. Uttar Pradesh is a north Indian state in which in 41% of women are illiterate according to the 2011 Census, and in which the TA+ program has reached more than 103,000 learners.[7]

Discussions with program staff and program participants indicated that, in almost all cases, women had to ask for permission from their spouses and in-laws before participating in TA+. For women who wanted to participate, an often-mentioned motivation was to not be illiterate and not to be looked down on. The ability to help children with homework and to not be cheated in transactions at stores were other important motivations. At the beginning of the program, staff of Development Alternatives (DA) spent substantial effort convincing local leaders, women and their families of the importance of female literacy. Indeed, 94% of learners had never attended school. Women were generally aware of computers, but had not handled them. In our sample, around 40% of the households in which the women were living had a mobile phone. All women were Hindi speakers. The characteristics of the women who participated in our study are further described in Section 5 below.

Our research was conducted in two phases corresponding to two different districts in Uttar Pradesh. The villages selected for the two phases differed somewhat. DA had previously worked in the villages of Phase 1 (though not with these women). As such the women who were signing up (and their families) were aware of the necessary time commitment and effort. Staff from DA also noted that there was a reinforcing dynamic from previous classes as neighbors of women who had completed the program did not want to be left behind. During Phase 2, DA was working in new villages, and staff noted that it was more difficult to motivate learners to sign up to the first round of classes within a village. Women were also more likely to sign up and then not participate in the program after being selected in the lottery.

4. Experimental Design

4.1 Sampling:

As the first step, working with DA, we compiled a list of illiterate learners in each village. Some women indicated that they were not interested in participating; consequently, they were not included in the lottery that selected learners nor in any of the analysis that follows. The interested women were grouped by the hamlet in which the classes would take place. The women were grouped by *jati* (caste) when DA thought respondents would be reluctant to join a class with learners from other castes. Each list of potential participants was entered in a public lottery to determine who would be assigned to the treatment group (i.e., enrolled in the DA program). In practice, the names of all potential participants were written on pieces of paper and placed in a tombola from which members of the public (usually children) drew the participants in the treatment group. After the treatment group was chosen, the remaining women were assigned numbers and placed on a waiting list. However, some women who had been selected for the treatment group did not participate, often because other household members did not permit the women to do so. These women were replaced by others, based on their rank on the waiting list.[8] Any woman on the lottery lists who was invited to participate in TA+ was considered a treatment respondent. The remaining women on the waiting list—who were not contacted—constitute the control group.[9] The women who formed the control group were invited to participate in TA+ in a subsequent round of classes.

4.2 Survey and lottery:

The sample for this paper is drawn from both two phases of the program. For each phase, the sequence of data collection was as follows: we first conducted a baseline survey, which was followed by the intervention (i.e., the TA+ program), which was then followed by an endline survey. The baseline survey consisted of questions on household composition, assets, and other socioeconomic characteristics of the women, which are described in detail in the following section. During the first phase in September 2013, 238 women from six villages in the Sant Ravidas Nagar district of Uttar Pradesh participated in our baseline survey. Using a public lottery,[10] 139 of these women were chosen for the treatment group and 99 were chosen for the control group.

Since the first phase of the program documented statistically significant learning effects (see the first panel of Appendix Tables A1 and A2), the research team thought that a larger study based on a larger number of villages would provide external validity to the Phase 1 results. Hence, a larger sample was drawn from 12 additional villages. Again, as happened in Phase 1, interactions with villagers indicated that the families of some illiterate women did not allow them to participate in the program, and some women simply did not want to.[11] In Phase 1 a lottery was conducted among only those women who had obtained consent from their families; however, in Phase 2, the lottery included all women who were interested, whether or not they had received consent from their families.

The second phase of the program was implemented between June and August 2014, after baseline surveys had been administered. As in the first phase, surveys were conducted at the end of the program as well. In total, an additional 479 women participated in the baseline survey for the second phase.[12] Of these, 264 were part of the treatment group and 215 were part of the control group.[13]

4.3 Intervention:

For Phase 1, TA+ was implemented during October and November 2013, before the endline survey in December 2013. The intervention spanned 26 days and was divided into three parts. Part 1 involved computer-based instruction for 100 minutes per day. The first 10 days were devoted to learning the Hindi alphabet (Devnagari script), and subsequent days of instruction were focused on decoding[14] and writing syllables and words. The program relies heavily on the use of mnemotechnic strategies, or "memory hooks," to help learners encode characters or syllables more effectively and maximize long-term retention. For example, each letter is linked with objects that start with that particular letter. These lessons were in Hindi, but an illustrative English example would be to show the letter *S* as a snake curled in an *S* shape. Each instructional session consisted of the following activities: a) 4 minutes of watching a video to facilitate the encoding[15] and recall of characters, syllables, and words; b) 12 minutes of work with large flash cards showing letters, characters, syllables, or words; c) 20 minutes of writing practice; d) 20 minutes of work with small flash cards to facilitate recall of letters, characters, syllables, or words that had been introduced; e) 10 minutes of identifying letters, characters, syllables, or words with the learners using the computer software; f) 20 minutes of writing practice; g) 10 minutes of quizzes and practice through peer learning; and h) 4 minutes for a follow-up video.

Part 2 spanned 10 days, with 100 minutes of instruction each day. This part of the program was primarily devoted to practice with reading (70 minutes) and writing (30 minutes). Part 3 lasted six days and focused on learning recovery. This part was designed to assist learners who were late in joining the program, had missed classes, or struggled with any part of the teaching material. Review days were interspersed during the course of the program, on the 6th, 10th, 20th, and 25th days. On the 26th day, all learners were assessed on their ability to recognize Hindi characters, write words, write phrases and sentences, and apply reading and writing skills beyond the program coverage. In our estimation exercise, we do not use make use of those assessments because they were not conducted by us.

TARA Ganit, the numeracy program, lasted 13 days, with each instructional session lasting 105 minutes. The learners' numeracy skills were assessed on the last day. Every session involved the following activities: a) 13 minutes on a story; b) 10 minutes of work with big flash cards; c) 12 minutes of writing practice; d) 10 minutes of work with small flash cards to recall numbers that had been introduced; e) 10 minutes of identifying numbers with the learners using the computer software; f) 10 minutes of instruction on writing numbers alphabetically; g) 10 minutes repeating the digits taught; and h) 30 minutes of writing practice. The 7th, 9th, 11th, and 13th days of the program were reserved for revising the numeracy learning. During this numeracy training, the learners were taught multiplication tables up to 10.[16]

4.4 Attrition, Participation, and Non-compliance

Appendix Table A3 shows the number of observations in our sample in both phases, separately for baseline and endline surveys. Of the 238 respondents from Phase 1, we interviewed 232 at the endline (with two control and four treatment observations not participating in the endline interview). Of the 479 respondents from Phase 2, we interviewed 430 at the endline (with 25 control and 26 treatment observations not participating in the endline interview).

Not all the respondents selected for treatment participated in the literacy program, and a very small number of the respondents in the control group participated. In the first phase, 127 of 135 women in the treatment group participated in the treatment (of which five attended only a few classes). In the control group, 10 women participated in some TA+ lessons, with six attending a few classes. In the second phase, 176 women of the 238 assigned to the treatment group participated in TA+, with 23 attending only a few lessons. Eight women from the control group participated, with two of these attending only a few lessons. In the estimation, we treat women who attended *any* classes as participants.

4.5 Tests

We administered a battery of literacy and numeracy tests to all the women (in the treatment and control groups) before and after the program. The literacy tests were developed, tried, and tested by Pratham[17] based on the model used in the DIBELS (Dynamic Indicators of Basic Early Literacy Skills) tests (Good, Kaminski, Simmons, & Kane'enui, 2001). The same model was used in the development of the Early Grade Reading Assessment (EGRA) and Early Grade Math Assessment (EGMA), which are now routinely used in international literacy and numeracy programs. Dubeck and Gove (2015) provide information on the history of EGRA development and the theoretical underpinning of the EGRA. They also provide references to other studies using EGRA. For example, Piper, Zuilkowski and Mugenda (2014) provide evidence from a randomized controlled trial on the effectiveness of on an early grade reading program in Kenya, where reading outcomes are measured using EGRA. Appendix B contains a more detailed description of the literacy and numeracy measures, although we outline the tasks here.

The literacy tests were timed tasks. The learners were given one minute to read 52 letters (Task 1), 63 syllables (Task 2), 52 words (Task 3), 48 nonwords (Task 4), and a 64-word Grade 1–level paragraph (Task 5). On Task 5, examiners marked as correct/wrong only those words read within a minute, but they allowed the learners to complete the paragraph after the first minute so that they could answer the comprehension questions. The test also included a number of "discontinuation rules" so the test would not progress to more difficult tasks if the learners could not achieve a minimum level on an earlier task.[18]

The math tests were not timed. Respondents were asked to complete eight tasks in Section I: (i) count three objects orally; (ii) recognize single-digit numbers ranging from 0 to 9; (iii) recognize 10 randomly selected two-digit numbers; (iv) count objects and circle the correct written number; (v) count objects and *write* the correct number; (vi) fill in the missing digit in two series; (vii) add two to three one- and two-digit numbers; and (viii) subtract one- and two-digit numbers. Based on the results of these eight tasks, we created an overall math score.[19]

5. Data and Sample Analysis

We present summary statistics for both the treatment and control groups, illustrating that the TA program targets a very disadvantaged population.

The process of selecting participants in the treatment group through a lottery was meant to create two groups with very similar observable and unobservable characteristics. A standard check to determine whether the randomization procedure was conducted properly involves investigating whether these two groups have similar observable characteristics, for example, demographic variables and baseline test scores. If one lottery had been held for all potential participants across all villages, then a simple t-test would have been appropriate. However, as Duflo, Glennerster, and Kremer (2007) note, one should include subgroup indicators (called *strata*) in the regression in this case.[20] This is also true for baseline balance tests. In our case, the strata fixed effects indicate the hamlet where the respondent lived or attended class.

[Table 1 about here]

Table 1 reports the summary statistics of individual- and household-level characteristics for the treatment and control groups, as well as whether the difference between the two groups is significant. Appendix Table A4 provides information on how the samples differed between Phases 1 and 2, as well as for the sample as a whole. Ninety-four percent of women in our sample had never attended school. Four percent belonged to the upper castes, 51 percent to the Other Backward Classes (OBCs), and the remainder to the Scheduled Caste or Scheduled Tribe (SC/ST) category.[21] The average age was 34 years. Approximately 90 percent of the women were married, 5.5 percent were unmarried, and 4 percent were widows. The average number of children in the household was 3.3. As a summary measure of household well-being, we computed scores for each household on Grameen Banks's Progress-out-of-Poverty Index (PPI) for India.[22] A PPI of 20 corresponds to a roughly 90 percent chance of being under the poverty line using international US$2/day/person poverty lines. The average PPI score of households in our sample was 25. Of the households, 30 percent had a Below Poverty Line (BPL) card.[23] The overall housing condition of our sample reflects the PPI index: 26 percent had access to electricity; approximately 57 percent resided in a brick house; only 4 percent had toilet facilities at home; 38 percent had access to a safe source of drinking water (i.e., access to private tap water or a community well); and 2 percent used a clean fuel source (i.e., a gas cylinder or electricity) for cooking. In terms of household assets, only 3 percent possessed any thermoware food-heating equipment (e.g., a thermos or casserole), 23 percent had either a TV or DVD player/VCD player/VCR, and 39 percent reported owning a personal mobile phone.

The third column of Table 1 which reports an estimate of the difference between the treatment and control groups, shows that the groups are balanced on virtually all characteristics.[24] The only difference is that the treatment group is 5 percentage points more likely to own a sewing machine, a difference that is statistically significant at the 5-percent level.

[Table 2 about here]

Table 2 gives an overview of the primary occupations of the women.[25] Approximately 76 percent of the women reported non-income-generating activities (household activities and chores) as their primary occupation. Classifying women using broad occupational divisions, 18.71 percent were involved in farm activities, 5.1 percent were involved in non-farm activities.

6. Impact Estimation

6.1 Empirical Framework

We first show that at the baseline the treatment and control groups were also similar in the outcomes of interest, namely, literacy and numeracy test scores. Columns 1 and 2 of Table 3 report the baseline means of literacy and numeracy outcomes for the treatment and control groups, respectively. Column 3 reports an estimate of the difference between the two. The table shows that the baseline (pre-treatment) difference in test scores between the two groups was not significant, with the exception of the ability to complete subtraction problems, as there is some evidence that the treatment group solved more problems ($p < 0.1$).

[Table 3 about here]

We then estimate a simple treatment effect model before implementing instrumental variables, difference-in-difference, and fixed effects estimations. Formally, we first estimate the following regression, where $testscore_{i,j}$ is the test score of individual *i* located in village *j*:

$$testscore_{ij} = \alpha + \beta_{ITT} ITT_{ij} + \mu_j + \varepsilon_{ij} \qquad (1)$$

ITT_{ij} is an indicator variable that takes a value of 1 if a respondent was assigned to the treatment group and a value of 0 otherwise. Because all the women did not necessarily comply with this assignment, the variable β_{ITT} captures the "intent-to-treat" effect, that is, the effect of being assigned to treatment. The variable μ_j is the specific sub-village (usually the hamlet) part of the error term (as mentioned above, our sample is stratified at the sub-village level), and ε_{ij} is the individual-specific part of the error term. All our standard errors account for clustering at the village level.[26]

Next, we include a set of individual- and household-specific control variables that convert equation (1) into the following regression equation:

$$testscore_{ij} = \alpha + \beta_{ITTC} ITT_{ij} + \gamma X_i + \mu_j + \varepsilon_{ij} \qquad (2)$$

The variable β_{ITT} captures the intent-to-treat effect for control variables. These control variables include age, marriage status, household assets, type of water the household has access to, type of cooking fuel used, material of the household dwelling, dummy variables for broad caste groups (SC, ST, OBC), PPI index, and whether the household has a BPL card.[27]

As noted above, not all households complied with the assignment to treatment and control groups. In particular, close to 20 percent of the women assigned to the treatment group did not participate in TA+. Therefore, intent-to-treat will be an underestimate of the program for those who participated in the treatment and we implement an instrumental variables strategy in which participation in TA+ is instrumented by assignment to the treatment group. In the first stage, we regress participation in TA+ on assignment to the treatment group and the other covariates used in the second stage.

$$TA_{ij} = \alpha + \beta_{Stage1} ITT_{ij} + \gamma_{Stage1} X_i + \mu_j + \varepsilon_{ij} \qquad (3a)$$

The instrumental variables (IV) approach then uses the predicted values instead of treatment assignment as an independent variable in the second stage.

$$testscore_{ij} = \alpha_{IV} + \beta_{IV} \widehat{TA}_{ij} + \gamma_{IV} X_i + \mu_j + \varepsilon_{ij} \qquad (3b)$$

The parameter β_{IV} captures the effect of TA+ for those induced into treatment by treatment assignment, and it is our preferred specification.[28] We correct the estimated standard errors for the two-stage procedure.

We implement a number of robustness specifications. The panel structure of the data allows the implementation of a difference-in-differences (DID) IV specification. Pooling the data from the first and second rounds, the equation to be estimated is as follows:

$$testscore_{ijt} = \alpha + \beta_{DDIV} \widehat{TA}_{it} + \delta_{DDIV} post_t + \theta ITT_i + \mu_j + \varepsilon_{ij} \qquad (4)$$

where $post_t$ indicates the information from the endline survey implemented post-intervention and ITT_i indicates whether a women was selected for the program in the lottery. β_{DDIV} is the parameter of interest on the participation variable instrumented by $post_t * ITT_i$.

Our last specification deals with fixed individual heterogeneity. We estimate the following fixed-effects instrumental variables specification where we allow for an individual specific fixed effect μ_i.

$$testscore_{ijt} = \alpha + \beta_{FEIV}\widehat{TA}_{it} + \delta_{FEIV}post_t + \mu_i + \varepsilon_{ij} \qquad (5)$$

6.2 Results

6.2.1 Reading Scores

In this section, we first present the impact estimates of TA+ on reading. We focus on whether these impact estimates are statistically significant. Then, in section 6.4, we turn to the interpretation of the magnitude of the effects.

Table 4 shows the main results of the impact of TA+ on literacy measures. The different columns contain results regarding the different measures of literacy in the order in which they were administered (which, with the exception of nonwords, corresponds roughly to the level of difficulty). The first two rows contain the mean and standard deviation of the control group using the endline data. The first panel[29] contains the estimation results for the simple difference estimate corresponding to equation (1). We find that the TA+ program improved literacy significantly across all dimensions of literacy. Compared with the control group, those assigned to the treatment group could read almost 9.5 more letters on average, 4.4 more syllables, and two to three more words depending on the difficulty level. These estimates are all significant at the 5-percent level. The results are robust to the inclusion of household and individual control variables, as seen in the second panel of Table 4.

The third panel presents the results of the instrumental variables specification. As expected, the coefficients are larger, reflecting the fact that the simple difference estimate does not consider that some individuals assigned to the treatment group did not participate. In particular, participation in TA+ increases the number of letters read in one minute by an average of 12.6 letters (compared with a control group mean of 3.4), the number of syllables by 5.7 (over a control group mean of 1.7), and the number of words read per minute by between 2.1 and 3.7 depending on the difficulty of the text selected.

[Table 4 about here]

6.2.2 Numeracy Score

Table 5 shows the results of TA+ on numeracy outcomes. Similar to Table 4, the first two rows contain the mean and standard deviation of the control group. With the exception of counting objects (where the mean at the baseline was already very high), TA+ has significant impacts on all mathematical tasks. Focusing on the absolute levels of these effects, learners in TA+ recognize, on average, about seven out of 10 written single-digit numbers and two out of 10 double-digit numbers, compared with three and one, respectively, for the control group. These results are robust to the inclusion of household and individual control variables and instrumental variables specifications as reported in the second and third panels of Table 5, respectively.

[Table 5 about here]

6.2.3 Robustness Checks

Tables 6 and 7 report the results from the DID and instrumental variables fixed-effects specifications with covariates. The results show that the estimated impacts are robust across a broad range of specifications and very similar to our previous estimate findings reported in Tables 4 and 5.

[Tables 6 and 7 are here]

We further tested the robustness of our findings by running phase-specific regressions using the specification used in the third panels of Tables 4 and 5. Appendix Tables A1 and A2 report the findings of these regressions on literacy and numeracy, respectively. The point estimates of the treatment effect on literacy outcomes are always positive and significantly different from zero with two exceptions: there is a positive but insignificant effect of TA+ on the number of words per minute read at the grade 1 and grade 2 levels during Phase 2. However, for the numeracy test, both phase-specific regressions provide evidence of significant program effects except for the two-digit number identification in Phase 2.

We notice that the results during the first phase are much larger than during the second phase, and the differences are particularly pronounced for the more advanced literacy skills. However, due to the limited number of villages, differences between the two are not statistically significant. We do not have a conclusive argument for why this is the case, although some differences between the two phases are noteworthy. First, the villages of the first phase were wealthier. Second, because the program had been implemented multiple times in these villages, the teachers had more experience. Third, teachers reported that during the second phase they did not need to expend much effort to motivate learners, perhaps because TA+ was known. Fourth, the post-program survey was implemented within two days of the end of the program, whereas the tests in Phase 2 were implemented later, on average.

Appendix Tables A5 and A6 show that the results are robust to the inclusion of baseline test scores.

For the main regressions up to now, we excluded learners who recognized more than 10 letters (of 52 in the Hindi alphabet) in the baseline test, treating them as outliers. The official cutoff for the program was five to six letters. In Table A7 in the Appendix, we reported the effect of TA+ by including those initial high baseline score samples and testing for whether the effects on these learners were different. Although the estimated impacts for high baseline learners are larger than for low baseline learners, only the difference for the words-per-minute outcome is statistically significant. We conclude that there is evidence that learners with only a small amount of knowledge at the baseline benefit more from the program. Given that we would consider most of these learners to be illiterate (in that they cannot put the knowledge of a few letters to much use), a case can be made to extend the eligibility criteria to include the barely literate in the literacy program. An argument can also be made that the difficulty of the program is tailored to learners with some background knowledge, and that DA could consider a preliminary phase to bring those who are completely illiterate to a minimum level of literacy before starting the program.

During our program implementation, we noticed that some learners attended only some of the classes and thus did not complete the program; however, in the previous instrumental-variables regressions (e.g., the third set of regressions of Tables 4 and 5), we treated them as if they had been full participants. For Appendix Table A8, we estimate a regression considering only those participants who completed the program. Our point estimates show that the estimated impact is noticeably higher for those who completed the program in both literacy and numeracy tests compared with any type of participants.

6.3 Long-term Results

DA organizes reading groups after the end of the TA+ program in order for the learning gains to persist. The group sessions include activities in reading, writing, and math.[30] About nine or 10 months after the endline (and 12 months after the baseline), we administered the same battery of tests to the women in the sample from Phase 2. Tables 8 and 9 contain the results using the

12-month literacy and numeracy test scores as dependent variables, where we implement the specifications from Table 4.[31] These point estimates of the 12-month effects can be compared to the estimates immediately after the program, which are presented in the second panels of Appendix Tables A1 and A2 (to compare IV estimates of Phase 2's immediate effect with the long-term effect). The results point to a slight loss in learning gains over this period.[32]

[Tables 8 and 9 are here]

6.4 Reading Scores in Perspective

One of the shortcomings of research on adult literacy in developing countries is that there is no benchmark for what literacy programs should specifically aim to achieve, though there have been attempts to change this (Abadzi, 2012)[33,34]. We are aware of no comparable data on reading outcomes of adult literacy learners in the Indian context that we could directly compare the learning outcomes of the TA+ adult literacy program with. Nonetheless, to provide some context for the meaning of our coefficient estimates, we first relate them to learning outcomes of children from the same context. Then we provide an attempt to contextualize the results by comparing them to an international benchmark. Both of these contextualizations are subject to substantial caveats that we note.

We compare the reading levels of the adult learners in our literacy program to those of schoolchildren in Uttar Pradesh and India using the 2013 Annual Status of Education Report (ASER, 2013). We do not in any way wish to imply that the learning progress in Indian primary schools should be the learning target for adult literacy programs. We are aware of the low levels of learning imparted in many rural primary schools, an issue that has become a topic in the media and in political discourse. Instead, we see this as an attempt to contextualize learning progress with other learners from the same social context.

Comparison with child learning outcomes in India

Table 10 compares the results of learners in TA+ to the ASER results from 2013, which include rural students in both public and private schools.

The first three columns refer to grade 1 and grade 2 students for Uttar Pradesh (columns 1 and 2) and all of India (column 3). The rows indicate the levels of the ASER test: children are classified at the "nothing," "letter," "word," "paragraph" (grade 1 text), and "story" (grade 2 text) levels based on defined performance criteria. For example, the inability to identify four out of five letters classifies the child at the "nothing" level.[35] From Table 10, we can see that 47 percent of children in grade 1 are classified under "nothing" in the all-India group (column 3).

Vagh (2012) compared the reading tests we use with the levels defined by the ASER tests for samples from the states of Bihar and Uttarakhand, two states that neighbor Uttar Pradesh. Using this comparison, we can benchmark our results against ASER levels by "converting" the results from the fluency test to ASER levels, which we do in the last four columns of Table 10.[36]

In particular, columns 4 through 7 present the computed proportions of control (columns 4 and 6) and treatment (columns 5 and 7) participants that fall into different ASER levels, based on the Bihar (columns 4 and 5) and Uttarakhand (columns 6 and 7) samples.

Comparing across columns, we note that TA+ moves learners away from the "nothing" category. TA+ learners perform approximately at the first-grade level of schoolchildren in Uttar Pradesh and in India as a whole.

[Table 10 about here]

There are a number of shortcomings of this comparison. First, we compare *learning increases* for TA+ with achievement *levels* for the schoolchildren. Thus, the extent to which children in Uttar Pradesh have a greater (or lesser) knowledge of letters before starting school relative to our control group will bias this comparison downward (respectively, upward) and change the assessment of the effectiveness of TA+. In all, 24.5 percent of the schoolchildren in Uttar Pradesh go to kindergarten, often as a requirement before attending private school (ASER, 2013).

Second, as noted above, Indian schools are heavily criticized for the lack of learning that takes place in them. Even though—if you take the comparison of adults and children at face value, which is a significant caveat—TARA Akshar is as effective as an average year of primary school in rural Uttar Pradesh in a much shorter period of time, this does not mean that the program has made most adults functionally literate.

Comparison with an international benchmark

In the absence of a standard profile of learning gains and meaningful targets in Indian adult literacy programs to which we could more convincingly compare TA+ learning outcomes, a second alternative is to compare the learning achievement to targets in international settings. Abadzi (2012) suggests that to understand about 80 percent of a text requires a words-per-minute score of at least 45. Although this comparison suffers from spanning different languages and alphabets (as noted also by Dubeck and Gove (2015)), it is nonetheless a useful benchmark. Turning to the learners in the TA+ program, these learners, on average, progressed from 1.7 to about 4.8 words per minute—substantially lower than the 45 words per minute recommended by Abadzi (2012). Learners with somewhat higher initial scores (and who are thus excluded from the main analysis) seem to benefit substantially more. While the control group learners are able to read on average 8.8 words per minute, TA+ participants with high initial scores are able to read about 21.3 words per minute more, for an average total of about 30 words per minute.

Overall, this suggests that TA+ can be effective. However, for the entirely illiterate learners (of which there are many), TA+ is not sufficient to bring them to a level that would be considered literate.

6.5 Numeracy Scores in Perspective

We are in a somewhat better position to contextualize the effectiveness of the TA+ program with respect to numeracy outcomes. To place the numeracy skills into perspective, we compare the effects of TA+ to those of a different program in India evaluated in a recent study (Banerji et al., 2015). This literacy program was conducted in 240 hamlets in two blocks (sub-districts) of the Purnia district in Bihar and two blocks of the Ajmer district in Rajasthan, and it was designed to determine whether a combined mother-and-child literacy program would improve child outcomes more than a mother literacy program alone or a child activity package alone. The "Mother Literacy" (ML) intervention provided daily literacy and numeracy classes. Volunteers were recruited from the community to teach classes for two hours per day over the course of 10 months. A version of Pratham's Read India methodology was modified to suit the interests of adults. (For more details, see Banerji et al., 2015.) In addition, a random sample of women also received a Child Home Activities and Materials Packet (CHAMP) to use with their children to improve their children's outcomes. In general, Banerji et al. (2015) found that the combined ML and CHAMP intervention yielded the largest effects.

The actual numeracy test used by Banerji et al. (2015) is different from ours in terms of the number and (perhaps) the difficulty of the test items within each task category (single-digit and double-digit number identification, single-digit addition, etc.). Therefore, we report a standardized measure of the effect of the program, called the "effect size" of both programs, as follows:

$$Effect\ size = \frac{mean\ of\ treatment - mean\ of\ control}{standard\ deviation\ of\ control}$$

We compare the effect sizes of the TA+ effects with the ones implied by Banerji et al. (2015)[37] in Table 11. We compare the outcome of TA+ relative to the ML program and the ML and CHAMP combination. Across all dimensions of numeracy, we find that the implied effect sizes of TA+ are larger than those of the program studied in Banerji et al. (2015). This suggests that the numeracy component of TA+ is more effective than traditional adult literacy programs.[38]

[Table 11 about here]

7. Conclusion

The world has seen substantial progress in access to primary education, with primary school enrollment rates for children in developing countries now reaching over 90 percent[39]. However, a large number of youth and young adults who are illiterate have been neglected in this expansion of access. Recognizing this problem in India, the NLM has now extended its remit to include youth and adolescents within its target group.

Nonetheless, adult, youth and adolescent illiteracy will be reduced through *effective* adult literacy programs only. In order to understand whether programs are effective, benchmarks need to be established against which programs can be measured using the same assessment tools. With benchmarks in place, effective adult literacy programs may be better able to advocate for funding. As Wagner and Kozma (2005) write, “One of the key impediments to expanding public and government support for adult literacy programmes has been the failure of those who support international adult literacy programmes to provide the type of reliable databases and impact evaluations typically utilized in other educational efforts.” The need for advocacy and the case for targeting resources to more effective programs is particularly salient in India—home to one-third of the world’s illiterate population.

In this paper, we investigate the impact of TARA Akshar plus (TA+), a computer-based adult literacy and numeracy program. Through this evaluation, we also provide some of the first rigorous evidence on the effectiveness of a computer-based adult literacy program. TA+ operates in the very disadvantaged, rural setting of villages in eastern Uttar Pradesh, where literacy rates among females are extremely low. We find that TA+ has statistically significant effects on literacy and numeracy skills, although the effects on reading are not large enough to make the vast majority of learners functionally literate. For the slightly less illiterate learners, the impacts are an order of magnitude larger, but they are still not large enough to make all learners read at a speed considered necessary for text comprehension.

These results, which occur over a very short period of time (two months), suggest two things. First, the differences between the entirely illiterate and the slightly less illiterate suggest that TA+ should not exclude the semi-illiterate from its courses, as these learners seem to benefit most profoundly. These may include school dropouts and semi-illiterate learners in less conservative and disadvantaged areas, as well as in urban settings. Indeed, it would be interesting to conduct further studies of TA+ effectiveness in such settings.

Second, entirely illiterate adults from as disadvantaged a background as the rural Uttar Pradesh villages we studied may need a precursor intervention to bring them up to a level where they can benefit. We hope that future iterations of TA+ and other adult literacy programs—computer assisted or not—make progress with these most challenging of learners.

References

Abadzi, H. (1994). What we know about the acquisition of adult literacy: Is there hope? World Bank Discussion Papers, No. 245. Washington, DC: World Bank. Retrieved from https://uta-ir.tdl.org/uta-ir/handle/10106/24626

Abadzi, H. (2003) Teaching adults to read better and faster: Results from an experiment in Burkina Faso. World Bank Policy Research Working Paper No. 3057. Retrieved from http://hdl.handle.net/10986/18181

Abadzi, H. (2012) Can adults become fluent readers in newly learned scripts? Education Research International, vol. 2012. doi:10.1155/2012/710785

Aker, J. C., Ksoll, C., & Lybbert, T. J. (2012) Can mobile phones improve learning? Evidence from a field experiment in Niger. American Economic Journal: Applied Economics 4(4), 94–120. doi:10.1257/app.4.4.94

Angrist, J. D., & Pischke, J.-S. (2010) The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. The Journal of Economic Perspectives 24.2, 3–30.

ASER Centre. (2013) Annual Status of Education Report (Rural). New Delhi, India: ASER Centre.

Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007) Remedying education: Evidence from randomized experiments in India. The Quarterly Journal of Economics 122(3), 1235–1264.

Banerji, R., Berry, J., & Shotland, M. (2017) The impact of mother literacy and participation programs on child learning: Evidence from a randomized evaluation in India. J-PAL Working Paper. Massachusets Institute of Technology, Cambridge, USA. Retrieved from https://www.povertyactionlab.org/evaluation/impact-mother-literacy-and-participation-programs-child-learning-india

Barrera-Osorio, F., & Linden, L. L. (2009) The use and misuse of computers in education: Evidence from a randomized experiment in Colombia. World Bank Policy Research Working Paper No. 4836, Impact Evaluation Series No. 29. DOI: http://dx.doi.org/10.1596/1813-9450-4836

Basu, K., A. Narayan and M. Ravallion (2001). "Is literacy shared within households? Theory and evidence for Bangladesh." Labour Economics 8(6): 649-665.

Berger, J. (2001) Effectiveness of computers in ALBE classrooms: An analytical review of the literature. Adult Basic Education 11(3), 162-183.

Beuermann, D. W., J. Cristia, S. Cueto, O. Malamud and Y. Cruz-Aguayo (2015). "One Laptop per Child at Home: Short-Term Impacts from a Randomized Experiment in Peru." American Economic Journal: Applied Economics **7**(2): 53-80.

Blunch, N.-H. (2013) Staying alive: Adult literacy programs and child mortality in rural Ghana. World Development 42, 114–126.

Blunch, N.-H., & Pörtner, C. C. (2011) Literacy, skills, and welfare: Effects of participation in adult literacy programs. Economic Development and Cultural Change 60(1), 17–66.

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2008) Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics 90(3), 414–27.

Card, D. E., Ibarrarán, P., & Villa, J. M. (2011) Building in an evaluation component for active labor market programs: A practitioner's guide. SPD Working Paper 1101, Inter-American Development Bank, Office of Strategic Planning and Development Effectiveness. Retrieved from https://ideas.repec.org/p/idb/spdwps/1101.html

Carron, G. (1990). The Functioning and Effects of the Kenya Literacy Program. African Studies Review, *33*(3), 97-120. doi:10.2307/524189

Cheung, A. & Slavin, R (2012). How features of educational technology applications affect student reading outcomes: A meta-analysis. Educational Research Review 7(3), 198–215.

Cheung, A. & Slavin, R (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. Educational Research Review 9, 88–113.

Chudgar, A. (2014) Mobile phone use and ownership among illiterate adults and its implications for adult literacy training. International Journal of Educational Development, 34(1), 20–29.

Deshpande, A. (2011) The grammar of caste: Economic discrimination in contemporary India. New Delhi: Oxford University Press.

Director of Census. (2011) Provisional results of 2011 census. Delhi & Mumbai, India: Offices of the State Directorate of Census.

Dubeck, M., Gove, A. (2015). "The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations." International Journal of Educational Development 40, 315-322.

Duflo, E., Glennerster, R., & Kremer, M. (2007) Using randomization in development economics research: A toolkit. Handbook of Development Economics 4, 3895–3962.

Gakidou, E., Cowling, K., Lozano, R., & Murray, C. J. L. (2010) Increased educational attainment and its effect on child mortality in 175 countries between 1970 and 2009: A systematic analysis. Lancet 376(9745), 959–974.

Gibson, J. (2001). "Literacy and Intrahousehold Externalities." World Development **29**(1): 155-166.

Good, R. H. III, Kaminski, R. A., Simmons, D., & Kame'enui, E. J. (2001) Using dynamic indicators of basic literacy skills (DIBELS) in an outcome driven model: Steps to reading outomes. OSSC Bulletin, 44(1), 1–24. Retrieved from https://eric.ed.gov/?id=ED453526

Haddad, L. (1999) The income earned by women: Impacts on welfare outcomes. Agricultural Economics 20(2), 135–141.

Hopkins, J., Levin, C., & Haddad, L. (1994) Women's income and household expenditure patterns: Gender or flow? Evidence from Niger. American Journal of Agricultural Economics 76(5), 1219–1225.

Iftekhar, S. N., & Hyeon, J. (2016) ICT based education for literacy in India: Vision beyond 2015. International Journal of Information and Education Technology vol. 6 no. 5, 388–393.

Imbens, G., King, G., McKenzie, D., & Ridder, G. (2009) On the Benefits of Stratification in Randomized Experiments. Unpublished manuscript, Department of Economics, Harvard University.

Kapur, S. & Murthi, M. (2011) Literacy in India . In K. Basu & A. Maertens (Eds.), The Oxford Companion to Economics in India (2nd ed., pp. 533-538). Oxford, England: Oxford University Press.

Lai, F., Zhang, L., Hu, X., Qu, Q., Shi, Y., Qiao, Y., Boswell, M., & Rozelle, S. (2013) Computer assisted learning as extracurricular tutor? Evidence from a randomised experiment in rural boarding schools in Shaanxi. Journal of Development Effectiveness 5(2), 208–231.

Maddox, B. (2007) Worlds apart? Ethnographic reflections on "effective literacy" and intrahousehold externalities. World Development 35(3), 532–541.

Mo, D., Swinnen, J., Zhang, L., Yi, H., Qu, Q., Boswell, M., & Rozelle, S. (2013) Can one-to-one computing narrow the digital divide and the educational gap in China? The case of Beijing migrant schools. World Development 46, 14–29.

Ortega, D. & Rodríguez, F. (2008) Freed from illiteracy? A closer look at Venezuela's Misión Robinson literacy campaign. Economic Development and Cultural Change 57(1), 1–30.

Oxenham, J. (2002) Skills and literacy training for better livelihoods: A review of approaches and experiences. African Region Human Development Working Paper Series, report no. 23957, Retrieved from https://openknowledge.worldbank.org/handle/10986/9767

Paxson, C., & Schady, N. (2007) Cognitive development among young children in Ecuador: The roles of wealth, health, and parenting. The Journal of Human Resources 42(1), 49–84.

Piper, B., Zuilkowski, S.S., & Mugenda, A. (2014) Improving reading outcomes in Kenya: first-year effects of the PRIMR Initiative. International Journal of Educational Development, 37, 11–21.

Senauer, B., & Garcia, M. (1991) Determinants of the nutrition and health status of preschool children: An analysis with longitudinal data. Economic Development and Cultural Change 39(2), 371–389.

Senauer, B., Garcia, M., & Jacinto, E. (1988) Determinants of the intrahousehold allocation of food in the rural Philippines. American Journal of Agricultural Economics 70(1), 170–80.

Strauss, J., & Thomas, D. (1995) Human resources: Empirical modeling of household and family decisions. Handbook of Development Economics. Amsterdam, The Netherlands: Elsevier. Retrieved from http://econpapers.repec.org/bookchap/eeedevchp/3-34.htm

Tamim, R., Bernard, R., Borokhovski, E., Abrami, P. & Schmid, R. (2011) What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. Review of Educational Research 81(1), 4-28. doi:10.3102/0034654310393361

Thomas, D. (1990) Intra-household resource allocation: An inferential approach. Journal of Human Resources 25(4), 635–664.

UNESCO Institute for Statistics. (2015) Adult and youth literacy fact sheet. Montreal, Canada.

Vagh, S. B. (2012) Validating the ASER testing tools: Comparisons with reading fluency measures and the Read India measures. Mimeo. New Delhi, India: ASER Centre.

Vagh, S.B., The Abdul Latif Jameel Poverty Action Lab (J-PAL), Pratham and ASER Centre (2012) Evaluating Read India: The development of tools for assessing Hindi reading and writing ability and math skills of rural Indian children in grades 1–5. Mimeo. New Delhi, India: ASER Centre.

Wagner, D., & Kozma, R. 2005. New technologies for literacy and adult education: A global perspective. Paris, France: UNESCO.

Wagner, D.A., Daswani, C.J., & Karnati, R. (2010). Technology and Mother-Tongue Literacy in Southern India: Impact Studies among Young Children and Out-of-School Youth. Information Technology and International Development, Vol. 6(4), 23-43.

White, H., & Masset, E. (2003) The importance of household size and composition in constructing poverty profiles: An illustration from Vietnam. Development and Change 34(1), 105–126.

Appendix B: Reading and Numeracy

For the TA+ outcome evaluation, literacy performance indicators were required. This set of indicators was complemented with indicators of numeracy performance, which often develop with literacy, and other cognitive indicators, which help in the interpretation of literacy and numeracy outcomes.

The specific tests we use were developed by the ASER Centre to study how well the ASER tests compare with EGRA/EGMA-type tests. We chose them for a variety of reasons.

- These measures were modeled after the EGRA/EGMA indicators developed by RTI International (an international nonprofit organization headquartered in North Carolina, USA) and the Dynamic Measurement Group (developers of the well-known DIBELS indicators, headquartered in Oregon, USA). These indicators have been tested extensively, successfully adapted in many languages, and used in many literacy/numeracy projects around the world.
- The comparisons with ASER tests are well documented, and they have demonstrated psychometric properties. The reliability and validity of these measures are now well established.
- The measures we have adopted have been used extensively in the Read India project. They thus permit us the comparison with other literacy/numeracy program evaluation studies.
- These tests are designed to assess the mastery of different segmental units of the Hindi language, including *akshar*, *barakhadi*, real words, invented words, and sentences. This feature makes it possible to differentiate the impact foci of the intervention program. This level of detail in the range of indicators is particularly useful for a) extracting a differentiated view of an intervention program impact and b) tweaking the intervention program's components to improve its effectiveness in the future.

Reading test:

The Fluency Battery was adapted by ASER from the Early Grade Reading Assessment (developed by Research Triangle International) and the Dynamic Indicators of Basic Early Literacy Skills (University of Oregon Center on Teaching and Learning). The material was extensively evaluated and piloted to ensure its grade and content appropriateness for the population of interest. Scores for the fluency reading subtests represent the number of units (*akshar*s/words/nonwords) read accurately in one minute, and scores for the reading comprehension subtest represent the number of questions answered correctly. The total administration time for the Fluency Battery is about 10 minutes.

The assessment of fluency is based on the premise that the ability to read with sufficient speed and accuracy, is important to read well and to comprehend text. Fluent decodings of letters, letter combinations, words in list form, and words in connected text are important and robust correlates of early reading ability and comprehension. The automaticity of these lower-level skills ensures that limited cognitive resources, such as attention and memory, can be freed and allocated to the higher-level skills of meaning making. Hence, fluency measures, which are orally administered tests, are widely used to assess children's early reading ability in English and several other languages.

As in the numeracy test, the test is structured to increase in difficulty. The objective is to gauge the comfort level of the women in the village with regard to recognizing different letters and words and reading them to form sentences. This test also includes two exit points. First, if a respondent fails to read a single word in the third round, the test stops. Second, if a respondent fails to read more than three sentences in the fifth round, the test stops.

Numeracy test

This test involves counting the number of a particular object and recognizing the numbers in numeric form. In addition, respondents are asked to perform simple mathematical observations, for example, single-digit recognition, double-digit recognition, addition, and subtraction. The test includes two exit points. First, if a respondent makes four errors in a row or identifies fewer than four digits correctly in the second round (recognizing single-digit numbers), the test stops there. Second, in part three (recognition of double-digit numbers), if the respondent incorrectly answers or fails to recognize four in a row, the question (not the test) is stopped and to the respondent is given the next question.

Appendix Table B1. Overview: Reading

Akshar (letter) reading fluency	Learners are shown a randomly arranged set of characters (*akshar*s) from the Hindi alphasyllabary and asked to sound them out. The score indicates the number of characters (*akshar*s) correctly sounded out in one minute. The test is stopped if the respondent is not able to read more than three letters in one minute.
Barakhadi (consonant-vowel syllable) reading fluency	Learners are shown a randomly arranged set of consonant-vowel (CV) *akshar* units and asked to decode them orally. The score indicates the number of *barakhadi* units decoded correctly in one minute.
Word reading fluency	Learners are shown a list of one- or two-syllable words and asked to read them aloud. The score indicates the number of words read correctly in one minute. The test is stopped if the respondent is not able to read a single word in one minute.
Nonword reading fluency	Learners are shown a list of one- or two-syllable invented words (or nonwords) and asked to read them aloud. The score indicates the number of nonwords read correctly in one minute.
Grade 1–level passage reading fluency	Learners are asked to read aloud passages comprising seven sentences and 64 words. The score indexes the number of words read correctly in one minute. The test is stopped if the respondent is not able to read more than three lines. Here the stop rule applies not to sentences but to lines.
Grade 2–level passage reading fluency	Learners are asked to read aloud passages comprising nine sentences and 94 words. The score indexes the number of words read correctly in one minute.
Grade 1–level comprehension questions	Learners are asked to answer four comprehension questions on each passage of the grade 1–level passage reading fluency subtest. The score is the number of questions answered correctly.
Grade 2–level comprehension questions	Learners are asked to answer two comprehension questions on each passage of the grade 2–level passage reading fluency subtest. The score is the number of questions answered correctly.

Appendix Table B2. Overview: Math Assessment

Math Assessment	
Oral counting	Learners are shown sets of pencils and asked to count the number in each set and state this verbally.
Number identification: one digit	Learners are shown one-digit numbers and asked to name them.
Number identification: two digits	Learners are shown two-digit numbers and asked to name them.
Counting: one-to-one correspondence	Learners are shown sets of objects and asked to count the number of objects in each set, then circle the correct number.
Counting: one-to-one correspondence with writing	Learners are shown sets of objects and asked to count the number in each set, then write the correct number.
Missing item	Learners are shown series of numbers with one number missing in each series and asked to write the missing number. The score indicates the number of correct responses.
Addition problems	Learners are given addition problems (e.g., “How much is 1 and 2 together?”). The score indicates the number of correct responses.
Subtraction problems	Learners are given subtraction problems (e.g., “How much is 1 and 2 together?”). The score indicates the number of correct responses.

[1] Development Alternatives (DA) is a social enterprise dedicated to sustainable development. For more information on DA, please visit http://www.devalt.org/.
[2] http://taraakshar.org/index.php/results/, accessed February 3, 2015.
[3] A number of evaluations of OLPC interventions in developing countries are, as of yet, unpublished
[4] World Development Indicators: http://data.worldbank.org/indicator/SE.PRM.NENR, accessed February 4, 2015.
[5] http://www.nlm.nic.in/nlmgoals_nlm.htm, accessed February 2, 2015.
[6] As one can see, achievement of all these goals at the same time is not straightforward, and it is not clear whether the program prioritizes any particular component of this multifaceted set of objectives.
[7] http://taraakshar.org/index.php/results/ (accessed September 30, 2015).
[8] We follow Card, Ibarrarán, and Villa (2011), who suggest this approach to ensure costly slots in labor market programs are filled while preserving the treatment and control group research design.
[9] To be exact, in some cases TA instructors could not make contact with an individual on the waiting list and thus skipped to the next participant. Our information on that is incomplete; we do, however, know who was the last

individual contacted on the waiting list. We take the conservative approach of assigning to the treatment group all women ranked earlier in priority than that last individual.

[10] Where a public lottery was not possible, the research team conducted a draw of names through a computer-based algorithm.

[11] Major reasons for nonparticipation that respondents stated during the survey were a) not interested in learning at this age (32 percent), b) having an inconvenience or a family responsibility at home, such as having a baby or older person to look after, being pregnant, or being unable to take time out from household chores (16 percent) and c) TA center is located at an inconvenient location (14 percent).

[12] For a few of the hamlets, the number of women was not large enough to support a control group. We dropped hamlets when there were fewer than three learners left for the control group.

[13] In the first set of villages in Phase 2, the staff from the NGO did not make as many attempts to recruit the women on the waiting list as in the second set of villages. Empirically, we account for this through the strata/hamlet fixed effects.

[14] Word decoding is the process of converting visual characters into speech sounds.

[15] Word encoding refers to the process of writing down words from an external or internal speech input.

[16] In Phase 2, TA+ spanned 10 additional days, which were dedicated to periodic tests (not additional days of instruction). However, in all our regressions, we adequately control for this difference between the two phases.

[17] http://www.asercentre.org (accessed January 25, 2015).

[18] The specific discontinuation rule was as follows: if a learner could not read more than three letters in a minute, the reading test was stopped at Task 1. If the learner could not read a single word in a minute, the reading test was stopped at Task 3. If the learner could not read more than three sentences without any time restrictions, the reading test was stopped at Task 5.

[19] Because piloting suggested extremely low math skills, we also asked students in Phase 1 to count up to 30 and in steps of 10 to 100 as the first task In the fieldwork. Because very few people could not count to 30, we dropped this part for Phase 2. We also implemented the counting in steps of 10 to 100 for Phase 2, but we do not present these results here because they are not part of the EGRA-type test. In Phase 2, we also administered a verbal math test with 15 questions but do not report these results for the same reason.

[20] Technically speaking, whether it is necessary or simply advisable to include strata fixed effects depends on whether the same proportion of participants is chosen in each lottery. Imbens, King, McKenzie and Ridder (2009) showed that including strata fixed effects for different lotteries generally lowers the estimated standard errors for the estimated coefficients and is thus advisable. When the lottery is conducted for separate subgroups separately and the number of participants chosen for the program as a proportion of all possible participants differs by subgroup, it is necessary to include strata fixed effects to avoid biased results.

[21] Scheduled Castes and Scheduled Tribes are "lower ranked" stigmatized and marginalized castes (*jatis*) and tribes, eligible for affirmative action. Other Backward Classes (OBCs) are castes and communities that are not stigmatized like the SCs but are low enough on various socioeconomic indicators to warrant additional affirmative action. (For more detail on the caste system and its economic consequences, see Deshpande, 2011.)

[22] For more information on the construction of the index, see http://www.progressoutofpoverty.org/ppi-construction.

[23] The BPL card is given to households deemed to be below the official Indian poverty line to identify disadvantaged households requiring government assistance.

[24] This estimate accounts for stratification and clustering (cf. Empirical Framework section for the specification).

[25] For this multiple-response question, we designated the "primary occupation" as the occupation/activity in which a respondent spent most of her time.

[26] More recently, Cameron, Gelbach, and Miller (2008) have argued that with a small number of clusters, it is preferable to base the standard errors on a *Wild cluster-bootstrap percentile-t procedure* and impose the null hypothesis as opposed to clustering. When we implement their suggestion with sub-village fixed effects, which we need to avoid biased estimates while clustering at the village level, the standard errors seem to be too *small.* We thus present the more conservative standard errors clustered at the village level.

[27] The list of control variables used in the regressions is shown in Table 1.

[28] Given that there is very little noncompliance on the part of the control group, this is very close to the treatment effect on the treated (ToT); see Angrist and Pishke (2010).

[29] By Panel, we refer to different sets of regression specifications reported in tables. If there are two sets of regressions in a table, we mentioned those as upper panel and lower panel. If there are three regressions, we refer to them as first, second and third panel, respectively.
[30] The reading club in Phase 1 was called *apnipathshala,* and it continued for six months (two hours daily) after the TA+ program was completed. It was led by an unpaid volunteer in each TA center from the same village. In Phase 2, the reading club was called *gyanchoupalik,* which also continued for six months (three-hour-long sessions daily) after the program. It was led by a paid teaching assistant who was given two days of training prior to the beginning of the reading clubs and further periodic training. In Phase 2, each session included 45 minutes of reading, 45 minutes of writing, 45 minutes of games, and 45 minutes on math.
[31] The full set is available upon request, but the qualitative results are the same across specifications.
[32] There is some contradictory evidence on whether the learning after nine to 10 months are larger, smaller or the same as during the program. The mean scores for the treatment group decline slightly as noted, but the difference between treatment and control groups widens slightly. Because the average test scores in the control group decline between the end of the program and nine months later, these results are difficult to interpret.
[33] https://www.academia.edu/6922710/Standards_and_criteria_for_courses_aimed_at_teaching_basic_literacy.
[34] Indeed, Abadzi's (2003) survey of literacy programs notes that each of these programs have different targets that they measure up against and different measurement tools.
[35] See www.asercentre.org for testing tools and the annual reports for test administration details.
[36] Vagh (2012) provides information on the number of letters read by students at the letter, syllable, word, and paragraph levels. We categorize our students into these categories, starting with the lowest; the average number of letters identified per minute for TA learners in that category corresponds to the average number of letters identified by children in the study by Vagh. For example, if children at the "nothing" level read 2.3 letters per minute on average in Vagh's study, then we assign an average score of 2.3 to the learners in our "nothing" category. We do the same for the upper levels. The last category is made up of learners who score above the word level but do not reach the paragraph level on average.
[37] Banerji et al. (2015) also present results for reading, but their test expands on the ASER tests and their test items are not close to being comparable to the results we present here. However, in terms of effect sizes for literacy outcomes in general, we find a similar ranking to that of the numeracy results. The results Banerji et al. (2015) report are in terms of the intent-to-treat effect. We make the numbers comparable by accounting for the fact that only 34 percent of their sample participated in the treatment. Assuming no spillovers and only one-sided compliance, the instrumental variables effect is about three times the intent-to-treat effect.
[38] Banerji et al. (2015) also compute a value for the standard deviation increase per 100 USD spent, which is 0.3 to 0.37. As per personal communication with Development Alternatives, the per-learner cost of TARA Akshar is 99 USD per learner for the program size in Uttar Pradesh. The increase in math scores is about 0.75 standard deviations, which translates into a 0.75 standard deviation increase per 100 USD. Because the focus of the Mother-Child literacy program is to raise child outcomes and there is a significant impact on children, this comparison of maternal outcomes is not a like-for-like comparison and should be interpreted with caution. Nonetheless, with Banerji et al. (2015) there are now two adult literacy programs that provide cost-effectiveness calculations.
[39]http://www.un.org/millenniumgoals/2015_MDG_Report/pdf/MDG%202015%20rev%20(July%201).pdf.