Improving Literacy in Rural India: Cellphone Games in an After-School Program

Matthew Kam, Anuj Kumar, Shirley Jain, Akhil Mathur, and John Canny

***Abstract*—Literacy is one of the great challenges in the developing world. But universal education is an unattainable dream for those children who lack access to quality educational resources such as well-prepared teachers and schools. Worse, many of them do not attend school regularly due to their need to work for the family in the agricultural fields or households. This work commitment puts formal education far out of their reach. On the other hand, educational games on cellphones hold the promise of making learning more accessible and enjoyable. In our project's 4th year, we reached a stage where we could implement a semester-long pilot on cellphone-based learning. The pilot study took the form of an after-school program in a village in India. This paper reports on this summative learning assessment. While we found learning benefits across the board, it seemed that more of the gains accrued to those children who were better equipped to take advantage of this opportunity. We conclude with future directions for designing educational games that target less well-prepared children in developing regions.**

***Index Terms*—cellphone, English as a Second Language (ESL), literacy, mobile game, pilot study**

I. Introduction

LITERACY is one of the great challenges in developing regions. Despite huge improvements in recent decades, literacy levels in many poor countries remain shockingly low. Even more challenging is the tension between regional and global "power" languages, that economic opportunities are often closed to those literate only in a regional language. For instance, India is a country with 22 regional and 2 national languages, i.e. Hindi and English. But English, together with computer skills, are the two most requested skills in surveys of poor parents [1]. English is a great economic enabler. It is the language of all professions and higher education, but also important for mid-level service jobs: retail, clerical, teaching, law enforcement, etc. that are the most common steps above menial labor. The value of English is widely recognized by ordinary Indians [2], and it is in fact the poorest citizens who are lobbying most strongly to expand English teaching.

English is thus the language of power in India associated with the middle and upper classes [3][4]. In other developing regions, it is another language such as Spanish, Mandarin, or French which is not native to most of the population. We believe that many of our lessons will apply to other languages although our focus is on English as a Second Language (ESL).

But the public school systems in developing regions face insurmountable difficulties. In India, for example, we were consistently unable to converse in English with those teachers responsible for teaching English in poor schools, where the overwhelming majority of children in the country struggle to learn. More important, public schooling is out of the reach of large numbers of children in rural areas and the urban slums who cannot attend school regularly, due to their need to work for the family in the agricultural fields or households [5].

At the same time, cellphones are increasingly adopted in the developing world, and an increasing fraction of these phones feature multimedia capabilities for gaming and photos. These devices are a promising vehicle for out-of-school learning to complement formal schooling. In particular, we believe that ESL learning games on cellphones present an opportunity to dramatically expand the reach of English learning, by making it possible to acquire ESL in out-of-school settings that can be more convenient than school. Games can make learning more engaging while incorporating good educational principles [6]. More important, a large-scale evaluation with urban slums children in India has shown significant learning benefits from games that target mathematics [7]. We believe that similar outcomes can be replicated with e-learning games that target literacy.

The challenge in evaluating any language learning project, however, is that language acquisition is a long-term process on the learner's part. Worse, with a novel technology solution that has yet to be institutionalized, there were tremendous logistical obstacles in running a pilot study over a non-trivial duration. After 3 years, in which we commenced with needs assessments and feasibility studies, followed by subsequent rounds of field testing interleaved with numerous iterations on our technology designs, we have established the necessary relationships with local partners for such an evaluation. This

Manuscript received September 22, 2008. This work was supported in part by the U.S. National Science Foundation under Grant 0326582, a Qualcomm BREW Wireless Reach award, and sponsorship-in-kind from Sony Creative Software.

Matthew Kam is an Assistant Professor in Carnegie Mellon University's Human-Computer Interaction Institute (phone: +1 412-268-9805; fax: +1 412-268-1266; email: mattkam@cs.cmu.edu). He was with the Berkeley Institute of Design and the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley when this work was carried out. John Canny is the founding director of BID and holds the Paul and Stacy Jacobs Distinguished Professorship of Engineering at UC Berkeley.

Shirley Jain participated in the pilot study as a curriculum developer and local supervisor.

Anuj Kumar and Akhil Mathur are with the Dhirubhai Ambani Institute of Information and Communication Technology, Gujarat, India.

paper describes the results from a semester-long pilot study – the longest so far in this project – which took place during the project's 4th year. The study involved 27 rural children who participated in an after-school program that we implemented in their village.

II. Related Work

Most work on technology-assisted language learning in the developing world does not explore the convenience that the cellphone's mobility offers. Banerjee et al. [7] report a large-scale evaluation with mathematics learning games on desktop computers, carried out over 2 years with urban slums children in India. Mitra et al. [8] describe a study in India with slums children over 5 months, which involved a "hole-in-the-wall" public computer installed with speech-to-text software. Dias et al. describe a computer-based tutor software for improving reading in Ghana [9]. Kothari's karaoke-like approach [10] targets native language – not second language – literacy using television.

With the cellphone's increasing ubiquity in Africa, Brown [11] argues that it is timely to envision a future where the cellphone plays a pivotal role in education in Africa. Kam et al. [12] describe how a set of ESL-learning cellphone games that targets children in rural India have undergone numerous iterations, based on successive, short-term formative studies. With the exception of Kam et al., the only education-related projects we know of in the developing world that leverage the cellphone are Islam et al. [13] in Bangladesh, and Librero et al. [14] in Mongolia and the Philippines. Both projects rely on Short Messaging Service and target university students, unlike our approach.

To date, Horowitz et al. [15] is the only study we know of that examines the cellphone for promoting literacy. However, this study took place in an industrialized country (USA), even though participants included households below the poverty line. In the study, Sesame Street videos that target the English alphabet were streamed to preschool children over cellphones throughout an 8-week period. Our paper therefore contributes to the literature as the first learning assessment on cellphone-based language instruction in developing regions.

He et al. [16] describe a 2-year randomized evaluation of a LeapPad-like device that supports custom software modules for English learning. This interactive system involves a paper book attached to a stylus and supports audio output. It lacks a visual display, unlike a cellphone, but overlaps with our goal of making literacy learning more accessible in the developing world using portable devices.

Among the learning technologies for developing countries, one of the most novel devices is the multiple-mice computer described in Pawar et al. [17]. It was intended for collocated learning by a group of children around each computer, with a mouse input device for each child. It has since been extended to distance learning in Moraveji et al. [18]. The evaluations in both papers were short-term, and underscored the difficulty in conducting a learning assessment for a novel educational technology over a substantial timeframe when it is not yet integrated into the everyday operations of a formal entity.

III. Overview of Pilot Study

The pilot study was carried out in collaboration with a non-government organization in North India under the terms of a Memorandum of Understanding. The study took the form of an after-school program, which we held during the afternoons at a private village school affiliated with this NGO. However, our goal was to investigate learning impacts that ESL learning games on cellphones have on lower-income rural children. As such, students who were already enrolled in this school were ineligible to participate in the study. Instead, we invited those parents who could not afford the fees for this private school – and hence sent their children to less expensive schools in the same area – to give consent for their children to participate.

In the after-school program, we ran three sessions per week, on average. Each session lasted two hours in the afternoon. Children from neighboring villages attended the after-school sessions after finishing their regular classes in the morning. In the after-school sessions, we loaned cellphones preloaded with ESL learning games to participants. The after-school program took place from late December 2007 to early April 2008, and spanned sessions on 38 days in total.

IV. Data Collection

As our preparation for this pilot study, we made two trips to India, i.e. once in the summer of 2007 to familiarize ourselves with the pilot location and end-user community, and a second time in December 2007 to kick-off the actual pilot. 4 local staff members were hired to run the after-school sessions on an everyday basis. 3 of them were engineering undergraduates in their last semester, while the last member had graduated a few years ago. We spent two weeks training them to run the after-school sessions and perform data collection, and continued to coordinate with them regularly via conference calls and emails after we left India.

We interviewed participants on their demographics such as their ages and the grades they were currently enrolled in in school. During the interviews, we also asked other questions, such as the number of cellphones that their households owned, what they currently and/or had previously used cellphones for, their television watching habits and frequency, as well as their parents' occupations. The questions on media and technology exposure were included because these variables were expected to impact participant ability to learn using cellphone games.

To ensure that each participant has the basic numeracy and ESL literacy to benefit from cellphone-based learning in the program, participants were required to pass a qualifying test, i.e. obtain at least 50% of the total score. The test required them to complete one-word blanks using English words about themselves, e.g. name, age, school, grade, etc. They were also

asked to fill in the missing letters in the alphabetic sequence, write numbers in the Arabic notation, match words with their pictures, spell the words for everyday objects, and describe a picture of a market scene with short sentences. The qualifying test was designed such that an average child in India with no learning disabilities who has finished 1st grade in a reputable urban school should obtain a perfect score on it.

By using the qualifying test as a screener, we ensured that participants were numerate. This was important because we had previously found it difficult to teach children to use the cellphone keypad's to play e-learning games when they were not familiar with the numbers from 0 to 9 in Arabic notation. Similarly, by ensuring that participants were familiar with the English alphabet, we could target a more advanced syllabus that went beyond the alphabet. We made this decision since Horowitz et al. [15] had already investigated the efficacy of cellphone-based learning for the English letters in the context of preschool children in the USA.

Since success in acquiring a second language is correlated with literacy in one's native language, we administered a test which evaluated the ability of the participants to read in Hindi. Every child was given a short passage that described a diet for promoting dental health. Each child was then asked to read the passage aloud so that we could observe his or her fluency and accuracy. These sessions were videotaped. After that, every participant was asked to write answers to written questions that tested his or her comprehension of the passage, in Hindi. We had designed this test such that an average child who has finished 3rd grade in a reputable urban school in India should obtain a perfect score on it.

Our primary method of assessment was to administer pre- and post-tests which evaluated participants on their ability to spell the common nouns that the curriculum for the pilot study targeted. Although the curriculum targeted other competencies such as listening comprehension and the recognition of written words, our assessment emphasized spelling, which as a recall task was cognitively more difficult than recognition tasks.

We maintained attendance records for the participants for every session. We also videotaped each session so as to have contextual data that could potentially account for their test performances. The video recordings captured the classroom proceedings, and individual participants' interactions with the games. The latter recordings captured participants' levels of engagement with the games as shown in their facial and body expressions. The pilot staff member who was responsible for videotaping the sessions tried to ensure that every participant was videotaped playing at least one level in the curriculum per day. The recordings were later transcribed and translated from Hindi to English.

Finally, for every session, we asked pilot personnel to write a report which summarized what happened in that session, as well as how well each participant interacted with the games. The latter not only covered usability and learning obstacles, but also included pilot staff's observations on the attitude and persistence that each child demonstrated towards learning.

V. Participants

Owing to the strong relationships that our NGO partner had built with the local community over more than a decade, we were able to generate a high level of support among parents in this community. In total, we obtained consent for 47 children to participate in the study. However, we needed to turn 16 of them away; 15 children did not pass the qualifying test while the 16th was attending private tuition for English. The latter represented a confounding variable. Of the 31 children whom we started the pilot with, 4 of them left the program mid-way. Reasons for attrition include time conflicts with private tuition (2 children) and disinterest in attending the sessions (another 2 children). From post-deployment interviews, we understand the latter was due to caste tensions between those 2 children, who belonged to the lower castes, and some upper-caste children in the program.

A. Demographics

The 27 children who participated in the study until it ended were aged 7 to 14 (mean = 11½ years) and belonged to grades 2 to 9 (mean = 6th grade). There were 11 boys and 16 girls. 5 children came from the upper castes while others belonged to the lower castes. The gender and caste breakdown seemed to mirror the demographics in the community. Every participant attended between 8 and 29 sessions (mean = 20) in the after-school program, broken down according to the following three functions:

- Cellphone training: 0 to 5 sessions (mean = 4) where we taught participants how to use the cellphones, perform alphanumeric input and play mobile games,
- ESL learning: 4 to 17 sessions (mean = 10) in which participants played ESL learning games on the cellphones, and
- Assessment: 4 to 7 sessions (mean = 6) for administrative tasks and data collection, e.g. demographics interviews and various tests.

In India, traditionally, only the upper castes owned land. As such, the upper castes earn their livelihood on the land or run small businesses, while lower castes graze their goats, work as daily-wage laborers or perform menial jobs in the homes of the upper castes. Land-owning and non-land-owning families told us that they earned up to 100,000 (US$2,500) and 50,000 rupees (US$1,250) respectively per year.

B. Hindi and English Baseline

26 of the 27 participants were enrolled in the same school, where Hindi is the medium of instruction. The last participant was a school dropout. Assuming regular school attendance, the typical participant would have taken classes on Hindi and English for 5½ and 3½ years respectively prior to the study.

We devised a grading rubric to evaluate each participant on the Hindi literacy test and qualifying test. On the former test, participants scored 7.9 out of 18 on average ($\sigma = 4.5$, $n = 19$). 2 participants turned in blank answer sheets. We observed the following problems in the submissions:

- Wrong answers due to poor comprehension of the questions (5% of the test-takers) or passage (53%), or responses that simply repeated the questions (32%)
- Spelling errors (16%)
- Grammatical errors, i.e. using the incorrect form of the verb for the subject's gender (21%), or the incorrect form of the noun for the subject's singularity vs. plurality (5%)
- Inability to phrase responses in complete sentences (5%)

TABLE I
BREAKDOWN OF PARTICIPANT PERFORMANCE ON QUALIFYING TEST

	Poor	Fair	Good
About myself	13% left blanks empty or filled them in Hindi	32% filled in blanks with at least 1 misspelling	55% filled in blanks with correct spellings for most questions
Alphabet	5% filled in less than 7 blanks in alphabetic sequence correctly	13% filled in ~11 blanks in alphabetic sequence correctly	82% filled in at least 12 out of 13 blanks in alphabetic sequence correctly
Word recognition	14% matched up to 2 out of 6 words with correct pictures	14% matched 4 out of 6 words with correct pictures	72% matched at least 5 out of 6 words with correct pictures
Spelling	41% spelt up to 2 out of 6 words correctly	18% spelt about 3 out of 6 words correctly	41% spelt at least 4 out of 6 words correctly
Picture description	18% made no attempt to answer this section	50% wrote answers as individual words, not full sentences	32% wrote intelligible answers in sentences

On the qualifying test, on average, participants scored 44.0 out of 50 ($\sigma = 5.5$, $n = 22$). Our grading rubric indicated how participants should be classified as "poor," "fair" and "good" on every section of the test. Table I gives the breakdown of how test-takers were distributed across categories for selected sections, and descriptions of the categories. In summary, the average participant had a good knowledge of the alphabet and a fair vocabulary of written words that she could read. On the other hand, she was weak in recalling and spelling everyday nouns, and even weaker in constructing complete sentences with these words. Despite the wide range in the ages of the participants, it appeared from their performance on the above tests that the variation in their English proficiency was much narrower. More specifically, we estimated that the average participant was comparable to an urban child in India who had taken between 1 and 2 years of English classes.

Notably, only 10 children (45%) could spell their names correctly in English on the qualifying test papers. We had a chance later to interview the teacher who taught them English in their school. She revealed that her pedagogical approach revolved around having students copy sentences from English textbooks into their notebooks. She felt that it was not worth putting in more effort to teach English since she believed she was underpaid.

C. Technology Baseline

Among the 27 participants, 25 of them came from families who owned at least one cellphone; 5 participants belonged to families that owned 2 phones each while 2 participants came from families which owned 3 phones each. The cellphone was usually used by the eldest male member in every family, and in fact, 3 boys aged 13-14 possessed their own cellphones. 8 participants came from families that owned cellphones with a color screen, as opposed to monochrome display. Two of the above cellphones – both of which belonged to 2 of the above 3 boys – contained built-in cameras.

Nonetheless, cellphone ownership and access were separate issues. Among the 25 children whose families owned at least one phone, 6 of them – 5 girls aged 7-11 and a boy aged 12 – were prohibited by their parents from using the phone, either entirely or most of the time. In general, in poorer families, it seemed that children were allowed to receive (free) incoming calls, but not play mobile games lest they drop the devices. On the other hand, in wealthier families, children were allowed to play on the phones. As such, although most participants were familiar with cellphones, it appeared that rural parents were more willing to entrust these relatively costly devices to their sons (vs. daughters). In total, 15 of the 27 participants (56%) reported that they had played cellphone games before.

VI. CURRICULUM AND GAME DESIGN

One of the major challenges with carrying out a pilot study over a non-trivial timeframe was that we needed to develop sufficient digital content that could last throughout its entire duration. We ensured that our syllabus was aligned with local ESL learning needs in India by recruiting a local ESL teacher as our curriculum developer. She had a decade's experience as an ESL teacher at a prestigious urban school, located in the same geographic region as the after-school program.

A. Curriculum Design

Given the above attendance rate, the ESL curriculum for the pilot was designed to be comparable to the amount of material that a qualified teacher could reasonably cover in 18 hours with rural children in a classroom. The syllabus was situated within the classroom theme, which participants could readily relate to. Concretely, the syllabus included:

- Common nouns that are found in the typical classroom, e.g. chair, table, door.
- Verbs that can be performed with the above nouns, e.g. sit, write, open, close.
- Sentence structures for constructing sentences out of the above nouns and verbs, e.g. "This is a __."
- Sentence structures for phrasing question-and-answer sequences with the above nouns and verbs, e.g. "What is this?", "Where is the __?"

The curriculum design took participant performance on the qualifying test into account. The curriculum was also based on our attempts to converse informally with participants, during which we learned they did not comprehend simple questions about themselves, did not know the English words for objects around them (e.g. in the classroom), and made grammatical errors. The curriculum therefore targeted the above syllabus in terms of listening comprehension, word recognition (of the written word), sentence construction and spelling.

B. *Game Design*

We designed a set of ESL learning games for the cellphone platform that targeted the above curriculum, and piloted them in the after-school program. Our designs drew on 3 resources, namely:

1) recurring patterns in state-of-the-art commercial software applications for language learning, which represented best practices that we reused to avoid reinventing the wheel,
2) traditional village games, which more closely matched the expectations and understandings that rural children have about games, in comparison to contemporary videogames that were largely Westernized, and
3) lessons from several previous rounds of field-testing and iterations with rural children elsewhere in India [12].

In this subsection, we walk the reader through a subset of the screen designs.

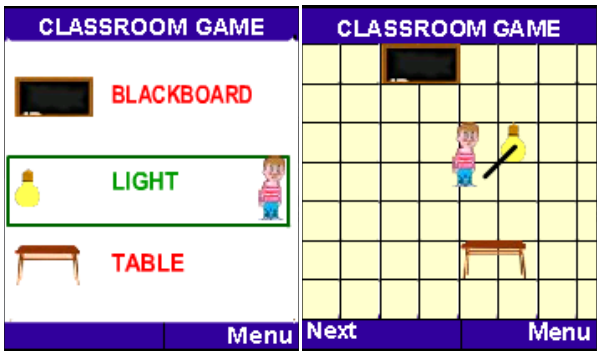

Fig. 1 and 2. Figure 1 introduces the English vocabulary for common nouns in the classroom. Word-picture association is a technique employed by many successful commercial language learning software. As the boy moves to each object, the software highlights the corresponding word in a different (green) color and says the word aloud. Figure 2 situates these objects in a classroom scene and builds on the earlier screen by demonstrating how to use the nouns in complete sentences. As the boy moves to each object, the software says the "This is a __" phase aloud for the corresponding object.

In earlier field studies, we observed that rural children did not readily associate a game with learning. It seemed that they viewed a game as an activity to be played purely for pleasure, and did not pay attention to the educational content embedded within game activities. On the other hand, when educational content such as English words and phrases were introduced in non-interactive screens separate from interactive game screens, the rural children appeared to grasp more intuitively that the software was trying to teach them those English words and phrases. Users subsequently paid more attention to the latter. Figures 1 to 4 show some screenshots in which we introduced words and phrases – both written and spoken – to the learners.

Fig. 3 and 4. Figures 3 and 4 introduce additional phrases that the nouns and verbs in the syllabus can be used in. Figure 3 teaches a phrase that associates the verb "sit" with the noun "chair." Figure 4 shows how to ask questions using the "Where" keyword. Abstract phrases and function words such as "where" are difficult to convey graphically. Hence, when they are taught for the first time, the software explains their meanings orally in Hindi.

The games tested players on their comprehension and recall of the words and phrases. For example, the game shown in Figure 5 says the word aloud for one of the objects displayed on screen. The player needs to identify the correct object and push it onto the area that is blinking blue. At the same time, he needs to avoid the balls thrown by the computer-controlled opponent. This game was an adaptation of Giti Phod, which was one of the traditional games that children play in Indian villages. In Giti Phod, players in a team have to arrange some objects (e.g. rocks) into a given configuration (e.g. a heap), while avoiding being hit by a ball thrown by members in the opposing team. In our experience, we have observed that rural children found it more intuitive to understand videogame rules when the designs of these videogames drew on the rules found in the traditional village games that they play everyday.

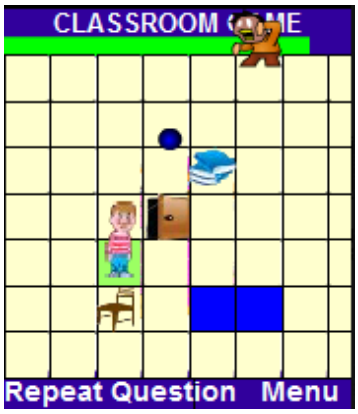

Fig. 5. A word-picture matching game which is an adaptation of one of the traditional village games that children in rural Indian play everyday.

Given that television has become a pervasive media among all economic classes in India, it only made sense to draw on popular culture in India to make our designs more appealing to children there. One of these sources is Sesame Street, which is a successful television program for young children that has local co-productions around the world – in both industrialized and developing countries. Its producers in India have found some of its localized characters to be popular with children in India, and we incorporated those characters into our designs

for teaching (Figure 4), quizzing (Figure 6) and congratulating (Figure 7) the user.

Fig. 6 and 7. Localized characters from the Indian production of Sesame Street tested the player on his ability to engage in question-and-answer style dialogues, and performed a victory dance for the player upon successful completion of each level in the game.

The activity that targets spelling skills is shown in Figure 8. The player is given an image (e.g. blackboard) and is required to spell the word corresponding to it. Some of the letters in the word are displayed, while blanks are shown for the remaining letters. The player moves between blanks with the arrow keys. Once he has filled in all the blanks and submitted his answer, the correct and wrong letters are displayed in green and red respectively. This feedback constitutes the first level of hints that we have designed to help the learner arrive at the correct spelling. If the player spells the word correctly, he proceeds to the next game. Otherwise, all blanks are cleared after a short pause and he is required to spell the same word again.

Fig. 8. The first level of hints in the spelling activity. After the player has tried to spell the word by filling in the blanks with letters, the correct and incorrect letters are shown in green and red respectively.

If the player is unsuccessful in spelling a word correctly after two attempts, the second level of hints (Figure 9) appears to provide him with additional learning support. Based on the blank that the cursor is currently located at, the game displays a set of possible letters for him to narrow down the choice of candidate letters.

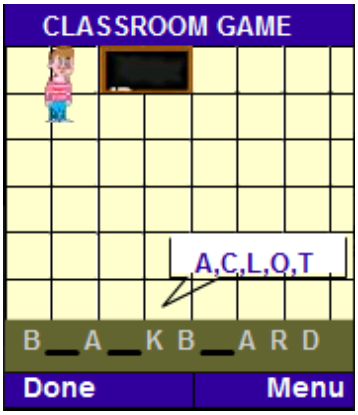

Fig. 9. The second level of hints in the spelling activity. For every blank, a set of possible letters are displayed to provide the player with some assistance, if he was unable to spell the word correctly after two attempts.

The curriculum is broken up into a total of 6 levels in the games. On every screen, the player can access a menu through a shortcut button. Among various options, this menu permits him to move to an earlier level in the curriculum to repeat the material, as well as to move to higher levels in the curriculum. The software was designed so as not to require airtime, which was expensive for most rural families. We implemented the games on Adobe's Flash Lite and Qualcomm's BREW (Binary Runtime Environment for Wireless) platforms. We piloted the games on Motorola's Razr V3m cellphone model, which has a fairly large screen.

VII. Pilot Sessions

In those sessions where participants were taught how to use the cellphone, they were shown how to move their sprites with the arrow buttons. They were also taught how to perform alphanumeric text entry, since most of them did not know this. Sprite movement and text input were essential skills for the games we designed. Pilot staff were therefore asked to write some simple, short sentences on the blackboard, and ensure that each participant demonstrated his ability to enter those sentences via text input.

Some other sessions focused on administrative tasks, such as the above tests and demographics interviews. We learned that a few participants had difficulty reading a small subset of the English alphabet despite having passed the qualifying test. We spent two sessions coaching them on those less-frequently encountered letters, so that they would be better prepared for the syllabus targeted in the pilot. Next, at least 8 children had seen the localized Sesame Street characters on television, but did not know their names. To help participants better relate to the characters, so that our games would appeal to them even more, we introduced the characters at the start of the semester. We also screened 3 localized episodes on separate occasions. These episodes were chosen such that they were educational but did not target English learning. Each episode lasted ½ hour, and we observed that participants enjoyed the humorous acts performed by the characters.

Fig. 10. In the after-school sessions, each participant was loaned a cellphone preloaded with English language learning games. Participants were taught how to start the games, and were asked to focus on learning English when playing the games on their own.

The remaining, and majority of, sessions focused on ESL learning. A 2-hour session was typically structured as follow: after an exchange of greetings, pilot staff took attendance and briefed participants on the learning objectives for that day. If new games were deployed that day, pilot staff explained and demonstrated how to play them to the participants in small groups. Each participant was then handed a cellphone to play the games on her own (Figure 10), and were told to focus on learning the English syllabus that the games covered. Children who were absent on previous sessions received help from pilot personnel in learning how to play those games that they were unfamiliar with. Pilot staff were limited to providing technical support; and were explicitly instructed not to teach English or communicate with participants in English. There was a short break of 10 to 15 minutes in the middle of each session. At the end of each session, pilot staff took back the phones so that they could charge their batteries overnight and download new games onto them. Each participants received a small packet of biscuits after every session.

VIII. Quantitative Results

On the pre- and post-tests, test-takers were awarded 1 point for each common noun in the syllabus that was spelt correctly.

A. Post-Test Gains

The mean pre-test score was 5.2 out of 18 ($\sigma = 3.3$, $n = 27$) while the mean post-test score was 8.4 out of 18 ($\sigma = 5.5$, $n = 24$). Participants exhibited significant post-test gains on a one-tailed t-test ($p = 0.007$). We present the frequency histograms for both scores in Figure 11. They illustrate that the score distribution had shifted toward the higher end of the spectrum after the deployment.

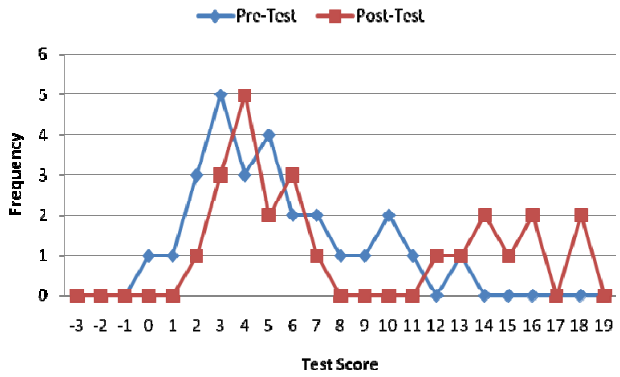

Fig. 11. Frequency histogram of participant scores on the pre- and post-tests.

The average post-test gains was 3.4 out of 18 ($\sigma = 3.3$, $n = 24$). The gains exhibited a fairly large variation, and ranged from -2 (two participants exhibited negative gains) to 9 out of 18. We present the frequency histogram for post-test gains in Figure 12.

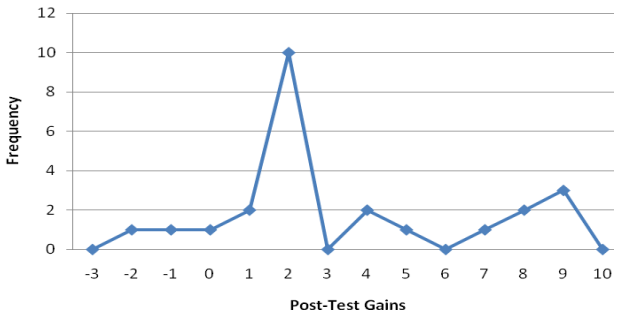

Fig. 12. Frequency histogram of participant post-test gains.

B. High-Gains vs. Low-Gains Learners

We sought to understand how participants' post-test gains were correlated with their demographics and performance on other tests. We also categorized participants into two groups, namely, "high-gains learners" and "low-gains learners," based on their post-test gains. A learner whose post-test gains exceeded the mean of 3.4 was categorized as a "high-gains learner," else he was classified as a "low-gains learner." In all, 9 participants were classified as high-gains learners while 15 participants were categorized as low-gains learners. 3 of the 27 participants could not be classified since they were absent on the day when the post-test was administered.

On a normalized scale, when the 27 participants were taken as one group, the average pre-test score was 29% whereas the average post-test score was 47%. The latter score did not seem high in absolute terms, i.e. on average, a participant could not spell over half of the common nouns targeted in the syllabus by the end of the intervention. However, once the participants had been classified, on a normalized scale, high-gains learners scored 41% (80%) on the pre-test (post-test) whereas low-gains learners scored 19% (27%) on the pre-test (post-test), on average. In other words, high-gains learners not only showed larger post-test gains but also appeared to have a higher mean

pre-test score at the start of the intervention. More important, post-test gains for both high-gains ($p < 0.001$) and low-gains learners ($p = 0.076$) were significant, with effect sizes of 0.54 and 2.24 respectively. That is, both categories of participants exhibited learning gains. (But average post-test gains for low-gains learners were only marginally significant, because of the 2 participants who obtained lower scores on the post-test compared to their pre-test.)

TABLE II
HIGH-GAINS VS. LOW-GAINS LEARNERS IN TERMS OF DEMOGRAPHICS

		Age (Years)	Grade Enrolled in School	Days Spent Learning ESL Games	Days Spent Playing ESL Games
Low-gains learners (n=15)	Mean	10.7	5th	3.9	10.5
	σ	1.8	1.9	1.4	4.2
	Min.	7	2nd	0	5
	Max.	14	9th	5	16
High-gains learners (n=9)	Mean	12.8	8th	3.8	10.2
	σ	1.2	0.9	0.8	5.2
	Min.	11	7th	2	4
	Max.	14	9th	5	17
Is difference between means significant?		Yes ($p = 0.002$)	Yes ($p < 0.001$)	No ($p = 0.4$)	No ($p = 0.4$)
Correlation with post-test gains (r)		0.45	0.61	0.11	0.10

In Tables II and III, we examined how high-gains learners may differ from low-gains learners in terms of demographics. On the whole, the high-gains learners did not appear to differ significantly from low-gains learners in terms of the number of days that they spent on learning how to play the cellphone-based games ($p = 0.4$) or actually playing the games to learn ESL ($p = 0.4$). Instead, high-gains learners belonged to higher ages ($p = 0.002$) and were enrolled in more advanced grades in school ($p < 0.001$). In fact, post-test gains exhibited high correlation with grade levels that participants were enrolled in school ($r = 0.61$) and medium correlation with age ($r = 0.45$).

TABLE III
HIGH-GAINS VS. LOW-GAINS LEARNERS IN TERMS OF DEMOGRAPHICS

	Sex	Caste	Media Exposure	Attitude*
Low-gains learners (n=15)	67% (33%) were females (males)	77% (23%) belonged to lower (upper) castes	73% have played games on cellphones prior to pilot; 60% (40%) watched less (equal to or more) than 1 hour of TV per day	24%, 38% & 38% were described as below average, average and above average learners respectively
High-gains learners (n=9)	44% (56%) were females (males)	88% (12%) belonged to lower (upper) castes	56% have played games on cellphones prior to pilot; 40% (60%) watched less (equal to or more) than 1 hour of TV per day	29% and 71% were described as below average and above average learners respectively

*The Attitude column is based on the observations that pilot personnel have on the seriousness and aptitude that participants exhibited as learners throughout the pilot. These qualitative comments were subsequently coded into the "below average", "average" and "above average" learner categories.

The proportions in Table III were presented for the sake of completeness. We were unable to perform any statistical tests on these proportions due to the small sample size, which for example did not satisfy the standard binomial requirement. We thus caution the reader against drawing firm conclusions from these statistics. However, when examining individual learners to identify surprising cases, we took the demographic variables in Table III into consideration. The analysis is deferred to the following section.

TABLE IV
HIGH-GAINS VS. LOW-GAINS LEARNERS IN TERMS OF TEST SCORES

		Qualifying Test (out of 50)	Qualifying Test, Spelling Section (out of 6 words)*	Hindi Literacy Test (out of 18)	Pre-Test (out of 18)	Post-Test (out of 18)
Low-gains learners (n=15)	Mean	42.9	1.2	6.3	3.5	4.8
	σ	2.9	1.1	4.2	2.3	2.4
	Min.	37	0	0	0	2
	Max.	46.5	4	14	10	12
High-gains learners (n=9)	Mean	47.1	3.4	12.0	7.4	14.4
	σ	1.8	1.6	1.7	3.1	3.6
	Min.	43.5	2	10.5	2	6
	Max.	49	6	14	13	18
Is difference significant?		Yes ($p < 0.001$)	Yes ($p = 0.001$)	Yes ($p < 0.001$)	Yes ($p = 0.003$)	Yes ($p < 0.001$)
Correlation with post-test gains (r)		0.57	0.70	0.45	0.46	0.86

*In this column, we present the number of words that participants spelt correctly on the spelling section of the qualifying test, out of a total of 6 words.

Table IV compares the high-gains and low-gains learners in terms of test scores. The former outperformed the latter on the Hindi test ($p < 0.001$). Next, we analyzed the qualifying test results at two levels, namely, the score for the entire test as well as the score on the spelling section. We found that high-gains learners outperformed low-gains learners on the entire test ($p < 0.001$) as well as on the spelling section ($p = 0.001$). High-gains learners also obtained higher scores on the pre-test ($p = 0.003$) and post-test ($p < 0.001$), vis-à-vis low-gains learners. In fact, participants' post-test gains exhibited a high degree of correlation with their qualifying test scores, for both the entire test ($r = 0.57$) and spelling section ($r = 0.70$). On the other hand, post-test gains had a lower correlation with Hindi literacy levels ($r = 0.45$) and pre-test scores ($r = 0.46$).

IX. QUALITATIVE RESULTS

The above quantitative results suggested that current levels of spelling proficiency and grades enrolled in school were the strongest predictors of success in learning how to spell new words through the cellphone-based games which we designed. Higher levels of Hindi literacy and academic preparation were also associated with higher post-test gains.

On the other hand, the number of sessions that participants had with the cellphone games – both for learning how to play the ESL learning games and learning ESL through the games

– were not associated with post-test achievements. Among the 24 children whom we have post-test gains data on, 5 of them were classified as high-gains learners despite having played the games on only 4 to 7 days (mean = 6 days). Conversely, 7 participants were classified as low-gains learners in spite of having played the e-learning games for 13 to 16 days (mean = 14.7 days). More important, pilot personnel described 6 of these 7 low-gains learners as "hardworking" or "serious" about learning ESL. Similarly, we were curious about how the two students who exhibited negative post-test gains, as well as the school dropout, had interacted with the cellphone-based games.

The above quantitative trends raise the following questions: How did some of the high-gains learners played the games such that they benefited despite lower attendance? In the case of some low-gains learners, why did they improve little on the post-test despite spending numerous days with the games and being perceived as diligent? In order to address such questions at the interaction design level, we turned to our video records and daily reports. Our hope was to recommend improvements to the technology designs and/or after-school setting.

A. Interaction Patterns with the Technology

At first glance, it seemed that participants needed to attend the after-school program for more days. Our video recordings showed that only 3 of the participants reached the last level in the curriculum by the last session in the program. This was a surprise. Given that the curriculum was designed for 18 hours of instruction, we expected an average attendance rate of 10 gameplay sessions to constitute enough time with the games. On examining the video recordings more closely, we saw that at least 8 participants were using the game menu to skip ahead to other levels whenever they were unable to spell the words in the current level correctly after a few attempts. (We note that the menu was not necessarily a negative feature. Among those 8 participants, at least 2 of them used the menu to skip those words that they already knew how to spell.)

We needed to understand why learners gave up on retrying the spelling activity for difficult words despite the hints in the spelling activity. On the whole, we observed 4 different levels of behavior associated with the spelling activity in the videos:

1) When students encountered a word that they could spell, they pressed the keypad buttons quickly and with ease to fill in the blanks for the missing letters.
2) When students saw a word that they did not know how to spell, some of them learned to spell it correctly with the help of the first level of hints.
3) Some of those students who failed to learn how to spell a word with the first level of hint eventually learned how to spell it correctly with the help of the second level of hints.
4) Other students never succeeded in learning how to spell certain words despite both levels of hints.

In general, we observed that high-gains learners succeeded in learning how to spell words after having seen their written forms displayed on earlier screens (i.e. such learners were able to spell those words correctly – without requiring any hints – on their first attempt in the spelling activity), or with only the first level of hints. It seemed that they did not require much scaffolding support from the software. In fact, from the video recordings of 9 high-gains learners, we saw that 5 (56%) and 1 (11%) of them depended on the first and second levels of hints respectively. In contrast, 12 (80%) and 8 (53%) out of the 15 low-gains learners who were videotaped relied on the first and second levels of hints respectively. It seemed that the low-gains learners, as compared to the high-gains learners, were less able to rectify their errors in filling in the blanks for the missing letters through the first level of hints, and required the second level of hints to attain the correct spellings.

Worse, the inability on the part of the low-gains learners to spell correctly with help from only the first level of hints made some of them visibly unhappy or bored when the second level of hints appeared. The reason for this distress was unclear. The learner could be frustrated that he was spending too much time to learn how to spell the word. Alternatively, on seeing the second level of hints show up, he could be demoralized that he had just been relegated to the ranks of the most inferior learners and needed the second level as a "crutch" in order to succeed.

Furthermore, some learners struggled despite both levels of hints. In the videos, two of them turned to their neighbors and asked for the correct letters, and/or to chat. In some cases, participants were embarrassed to ask their neighbors for help again after so soon, and hence used the menu to skip to other levels in the games.

More important, we observed that participants – especially among the low-gains learners – may be able to spell the words in the spelling activity, but were not able to spell the same words on the post-test. We offer two plausible explanations. Firstly, some children may have learned to spell the words by their last session in the program, but had forgotten their spellings between the last session and the post-test. Secondly, some participants never learned to spell the words in their entirety, since the spelling activity only involved filling a few blanks and did not require the learner to spell the entire word. Nowhere in the video recordings did we observe any child struggling with usability problems.

X. Conclusion

Our reactions to the results of the learning assessment were mixed. In an underdeveloped region where rural children did not have access to quality English instruction in their regular school or elsewhere, we were excited to see the participants -- both high-gains and low-gains learners – in the after-school program exhibit statistically significant post-test gains that could be reasonably attributed to our cellphone-based English learning games. On the other hand, the learning benefits were uneven among participants. This could be a cause for concern.

To begin with, high-gains learners outperformed low-gains participants on the pre-test, qualifying test and Hindi literacy test. In fact, participants' post-test gains appeared to be highly correlated with their existing levels of spelling proficiency (as

measured by their performance on the spelling section of the qualifying test) and the grades in school that they are currently enrolled in. This observation suggests that those rural children with a stronger academic foundation are the same children who are most well positioned to take advantage of the benefits that cellphone-based learning confers.

Our results are consistent with the outcomes of a study with rural and urban low-income children in India described in He et al. [16]. This study showed that weaker students benefited more from a teacher-directed pedagogical intervention, while stronger students benefited more from a self-paced, machine-based approach to English learning. These results should not, however, be interpreted to mean that we rule out technology-augmented learning completely in the context of low-income children. Horowitz et al [15] reported a study on videos for learning the English alphabet streamed over cellphones. In this study, a greater proportion of lower-income parents, vis-à-vis their higher-income counterparts, perceived the videos to have improved their children's knowledge of the alphabet.

In the face of the above overwhelming odds, what can we do to promote more equitable educational opportunities in the developing world? One possible – and perhaps cautiously optimistic – interpretation of the above results is that future research needs to be directed at understanding how e-learning software can provide more scaffolding support for those rural children who have less academic preparation. As an example, the spelling activity needs to be redesigned such that the learner is guided to spell the entire word eventually. With this redesign, however, gameplay becomes prolonged and can potentially increase player frustration, as we have witnessed above. One remedy is to have e-learning games track learner performance, so that the software can be adaptive in skipping stages that are similar to those that the player has previously performed well in. Another implication for instructional design, which calls for additional investigation, is scaffolds such as hints that are less conspicuous, so that their appearing on screen does not diminish the learner's sense of self-esteem or achievement.

Unfortunately, adaptive educational applications require the application state to be stored and retrieved on the same mobile device. From the logistics standpoint, this requirement is more difficult to implement in developing regions since it is harder to ensure that the same learner uses the same phone – which stores his performance from prior session(s) – across sessions. For instance, in an after-school program where attendance fluctuates from session to session, it would be prudent to keep a shared pool of cellphones, such that children who show up for the day's session can draw from. In these circumstances in which it is not possible to reserve a cellphone for each child, a wireless networked mechanism for synchronizing application state across all cellphones may be necessary. This, and other issues that we have raised above, require further investigation for cellphone-based literacy learning to be more effective in targeting less academically advanced rural children.

As we think more widely beyond the cellphone to consider it as a component in the broader learning environment, since the after-school program is a model that is readily replicable, we encourage the reader to adopt and experiment with the lessons from this paper. Our results suggest that the cellphone – which remains a relatively scarce resource in the developing world – is most effectively utilized in an after-school program that targets more advanced children. This restriction may be a necessarily evil until we gain a deeper understanding of how to design instructional scaffolds for less well-prepared rural children.

Next, children's tendency to seek help from their neighbors can be channeled productively if the latter are taught to offer help appropriately (e.g. instead of only telling their neighbors the correct spelling, help them to associate and remember the correct spelling). Such peer coaching strategies are especially crucial since cooperative group learning is unfamiliar to many rural children, whose schools (if they attend one) are more likely to implement rote learning. Alternatively, such an after-school program can hire facilitators to provide academically less prepared learners with similar coaching.

ACKNOWLEDGMENT

We thank Suraksha (Urvashi Sahni and Shalini Mathur) and Sesame Workshop India for collaborating with us on this pilot study. Lauren Bailey and Anuj Tewari provided much-needed administrative assistance, while Aman Anand and Siddhartha Lal provided technical assistance. We are especially indebted to Mehnaaz Abidi, Jatin Chaudhary, Neelima Purwar, Gautam Singh and Kavish Sinha for providing ground support. Lastly, we thank the parents and their children for their participation.

REFERENCES

[1] J. Pal, M.Lakshmanan, and K. Toyama, "'My child will be respected:' Parental perspectives on computers in rural India," in *Proc. of 2nd IEEE/ACM International Conference on Information and Communication Technologies and Development*, Bangalore, India, Dec. 2007.

[2] S. Shukla, "From pre-colonial to post-colonial: Educational transitions in southern Asia." in *Economic and Political Weekly*, 31(22), 1996, pp. 1344-49.

[3] D. Faust, and R. Nagar, "Politics of development in postcolonial India: English-medium education and social fracturing," in *Economic and Political Weekly*, India, July 28, 2001.

[4] M. P. Kishwar, "Deprivations's real language," in *The Indian Express*, September 15, 2005. Available: http://www.indianexpress.com/printerFriendly/12662.html.

[5] Azim Premji Foundation, "The social context of elementary education in rural India," 2004. Available: http://www.azimpremjifoundation.org/downloads/TheSocialContextofElementaryEductaioninRuralIndia.pdf

[6] J. P. Gee, *What Video Games Have to Teach Us About Learning and Literacy*, Palgrave Macmillan, 2004.

[7] A. Banerjee, S. Cole, E. Duflo, and L. Lindon, "Remedying education: Evidence from two randomized experiments in India," *NBER Working Paper No. 11904*, December 2005.

[8] S. Mitra, J. Tooley, P. Inamdar, and P. Dixon, "Improving English pronunciation: an automated instructional approach," in *Journal of Information Technologies and International Development*, Vol. 1, No. 1, Fall 2003, pp. 75-84.

[9] M. B. Dias, G. A. Mills-Tettey, and J. Mertz, "The TechBridgeWorld initiative: Broadening perspectives in computing technology education

and research," in *Proc. of the international symposium on Women and ICT: Creating Global Transformation*, ACM Press, 2005.

[10] B. Kothari, A. Pandey, and A. R. Chudgar, "Reading out of the 'idiot box': Same-language subtitling on television in India," in *Journal of Information Technologies and International Development*, Vol. 2, No. 1, Fall 2004, pp. 23-44.

[11] T. H. Brown, "The role of m-learning in the future of e-learning in Africa?," in *Proc. of 21st ICDE world conference*, Jun. 2003.

[12] M. Kam, A. Agarwal, A. Kumar, S. Lal, A. Mathur, A. Tewari, and J. Canny, "Designing e-learning games for rural children in India: A format for balancing learning with fun," in *Proc. of ACM international conference on Designing Interactive Systems*, Cape Town, South Africa, Feb. 2008.

[13] Y. M. Islam, Z. Rahman, S. S.Razzaq, M. A. Sayed, and S. Zaman, "Effect of feedback during lecture style delivery both in a face-to-face classroom and during a distance education television session in a developing country like Bangladesh without the use of Internet," in *Proc. of 6th IEEE international conference on Advanced Learning Technologies*, pp. 469-471, 2006.

[14] F. Librero, A. J. Ramos, A. I. Ranga, J. Triñona, and D. Lambert, "Uses of the cell phone for education in the Philippines and Mongolia," in *Distance Education*, vol. 28, no. 2, pp. 231-244, Aug. 2007.

[15] J. E. Horowitz, L. D. Sosenko, J. L. S. Hoffman, J. Ziobrowski, A. Tafoya, A. Haagenson, and S. Hahn, "Evaluation of the PBS Ready to Learn cell phone study: Learning letters with Elmo," reported prepared by WestEd, Sep. 2006.

[16] F. He, L. L. Linden, and M. MacLeod, "How to teach English in India: Testing the relative productivity of instruction methods within the Pratham English language education program," working paper, Jul. 1, 2008.

[17] U.S. Pawar, J. Pal, R. Gupta, and K. Toyama, "Multiple mice for retention tasks in disadvantaged schools," in *Proc. of ACM conference on Human Factors in Computing Systems*, San Jose, CA, Apr. 2007.

[18] N. Moraveji, T. Kim, J. Ge, U. S. Pawar, K. Inkpen, and K. Mulcahy, "Mischief: Supporting remote teaching in developing regions," *Systems*, pp. 753-762, Florence, Italy, Apr. 2008.