Mobilizing Education: Evaluation of a Mobile Learning Tool in a Low-Income School

Vanessa Frias-Martinez
Telefonica Research
Madrid, Spain
vanessa@tid.es

Jesus Virseda
Telefonica Research
Madrid, Spain
jvjerez@tid.es

Aldo Gomero
Pontificia Universidad Catolica
Lima, Peru
aldo.gos@gmail.com

ABSTRACT

The pervasiveness of feature phones in emerging economies has contributed to the advent of mobile learning applications for low-income populations. However, many of these tools lack the proper evaluation required to understand their educational impact. In this paper, we extend the state of the art by presenting the evaluation of a game-based mobile learning tool in both formal and informal settings at a low-income school in Lima, Peru. We show that *EducaMovil* improves knowledge acquisition in the formal environment of a classroom. In addition, use of the tool in more informal settings such as school breaks enhances the level of knowledge, as long as there is continuous engagement over time. We also demonstrate that *EducaMovil* can be used as a paperless complement to homework. Finally, we provide teachers with a set of guidelines for a successful deployment of *EducaMovil* at their schools.

Author Keywords

mobile games; low-income schools; learning gains

ACM Classification Keywords

K.3.1 Computers and Education: Computer Uses in Education

INTRODUCTION

Mobile devices have become an integral part of people's lives and a pervasive platform that we carry around at almost all times. In emerging economies, cell phone penetration rates are very high, frequently leapfrogging landlines and broadband access (ITU 2010). Such pervasiveness has contributed to the advent of various mobile-based services for low-income populations in areas like mobile education, mobile agriculture or mobile health. At its core, mobile services benefit from the familiarity that cell phone users have with the platform, which typically implies a small learning curve. Research in mobile education has covered a broad range of learning environments, from providing educational content as a complement to the formal setting of a classroom [4]; to enhancing learning experiences in informal environments like museums or distance learning [23, 25]. The motivations for using cell phones as learning tools are varied including the improvement of student access to learning materials, offering flexibility to students, reducing the costs of using more expensive technologies (like PCs), or to explore collaborative and individual learning [20]. In the specific case of emerging economies and low-income schools, learning environments become much more challenging. In general, these schools have limited budgets and scarce educational resources which often times have an impact on students' motivation and their performance. Given the high ownership rates of cell phones across the student population, mobile learning has been proposed as an affordable solution to complement formal schooling while keeping students engaged and motivated in the classroom. Additionally, the learn anywhere-anytime paradign also proposes the access to educational contents in informal settings e.g., while commuting or waiting for the public transport [12]. This is specially important in emerging economies where children might benefit from the access to educational contents in their downtime. However, there exists an important debate on whether mobile learning technologies constitute an adequate solution for students in low-income schools [21, 19]. Our research attempts to throw light onto this issue by providing a thorough evaluation of a mobile learning tool in a low-income school.

A large collection of mobile learning tools propose the use of smartphones or high-end phones to deliver educational content [25, 26]. Nevertheless, the availability of mobile learning tools for feature phones, which are the most common platform in emerging economies, is still very limited [22]. Most of the existing learning tools for feature phones deploy SMS- or Java-based solutions to provide school students with educational contents often presented as games [2, 10]. Kam *et al.* have shown that various mobile learning games (*MILLEE*) help rural kids to learn English in after-school programs in India [8]. Similarly, the *DrMath* project has offered Math tutors and learning content to hundreds of kids in South African schools [3]. Although these mobile games have shown promising results, they are typically criticized by their cost as well as by the lack of involvement of local teachers in the creation of contents. To overcome these issues, recent work like *EducaMovil* or the *English Literacy Tool* have advanced the state of the art introducing mobile learning games where the educational contents are created and automatically compiled into a game by the teachers themselves [11, 1]. These approaches have the advantage of highly re-

MobileHCI'12, September 21–24, 2012, San Francisco, CA, USA.

ducing educational costs while enhancing localized content authoring, both critical for low-income schools in emerging economies. Additionally, by involving teachers in the mobile learning process, it allows for a continuous and sustainable adaptation of the game to the curricular activities of the classroom. For these reasons, our research focuses on the evaluation of the educational impact of *EducaMovil*– a game-based mobile learning tool that is both low-cost and sustainable– at a low-income school in Lima, Peru.

Understanding the educational impact of a mobile learning tool in the long term is critical for its future deployments. Teachers need to evaluate whether such tools are convenient or not for their classrooms before involving themselves and their students in mobile learning programs. However, many of the mobile learning tools lack the proper evaluation required to understand their educational impact. Although there exist some studies on the impact of mobile learning tools in low-income schools, most of them either carry out only qualitative evaluations [15, 24]; or analyze the tool only in formal or informal settings but never both [16, 3]; or involve small numbers of students [4, 8]. Our research extends the state of the art by presenting a thorough quantitative and qualitative evaluation of the educational impact of *EducaMovil* when used in both formal and informal settings.

Our impact evaluation focuses on three important research questions: (i) the impact of *EducaMovil* on the students' knowledge acquisition *i.e., do students using the tool improve their level of knowledge more than students not using it? and, if so, under which formal or informal learning settings?*; (ii) the impact of the student's prior level of knowledge on her learning gains *i.e., do students need a minimum level of prior knowledge in order to benefit from the tool?*; and (iii) the relationship between game performance and knowledge acquisition *i.e., can teachers use EducaMovil as a proxy to measure a student's level of knowledge thus offering a paperless complement to student homework?* By answering these questions, we expect to provide outcomes and insights about the educational impact of deploying a game-based mobile learning tool at a low-income school. Our final objective is to propose a set of guidelines that will help teachers identify the ideal set up for the implementation of a successful *EducaMovil*learning program at their schools. Finally, although our findings are solely based on analyses for one specific school, we believe that teachers from other low-income schools with similar socio-economic and curricular conditions could potentially extend our results to their schools and decide on the usefulness of *EducaMovil* for their students. In the rest of the paper we present the related work; describe the low-income school and the mobile learning tool used in our evaluation; explain our research questions and methodology; describe our main research results and propose a set of guidelines for teachers willing to deploy *EducaMovil* at their schools. We finish with conclusions and future work.

RELATED WORK

Mobile Games

In recent years, the world has experienced an explosion in the amount of mobile learning tools that have been created for children and adult education in both emerging and developed economies. Game-based learning tools are present in different curricular areas such as: language learning [8] [18] or mathematics among others [7, 3, 9]. Kam *et al.* explored the use of mobile learning games to improve English literacy among Indian children [8]. The results of their pilot study with 27 students, showed that mobile phones have the potential to improve English literacy. However, the results were uneven among the participants: the students that had a higher English level at the beginning of the pilot benefited more from the use of the mobile learning application. In order to solve this discrepancy, the authors proposed as a solution the creation of adaptive mobile learning games that offer personalized contents depending on the students' knowledge. Similar research was presented in Tian *et al.* who studied how culturally inspired mobile learning games could help to improve Chinese literacy [18]. The research analyzes different traditional Chinese games, and uses them as a motivation to design cell phone games that follow similar rules. The authors show that culturally inspired games have more potential than western games to attract the interest of young students in China.

In the area of mathematics skills, a couple of interesting projects are *MobileMath* [7] and *DrMath* [3]. Although none of them rely solely on games for delivering educational content, the games are an important part of their educational models. *MobileMath* is an architecture that provides games, lessons, tutorials, examples and quizzes as separate entities for students to practice concepts in algebra. On the other hand, *DrMath* is a system that allows children to consult a tutor, participate in competitions, or to play single or multiple user text adventure games in the area of arithmetics. There exist other game-based mobile learning tools, like *EducaMovil* or *English Literacy Tool*, that provide more general frameworks to create quizzes for any curricular subject [11, 1]. These tools allow teachers to create collections of quizzes for different subjects, which enhances the creation of localized content authoring adaptable to the curricular activities of the classroom. Additionally, these tools do not require large budgets or engineering knowledge because they automatically embed the educational contents into the games while reducing the costs of creating educational games. Given that sustainability and low cost are critical variables for low-income schools, we focus our research on analyzing the educational impact of using *EducaMovil* at a low-resource school in Lima, Peru.

Impact Evaluation

There exists a large variety of methodologies to evaluate the educational impact of a mobile learning tool [15]. Price *et al.* proposed a research framework that grouped all methodologies into three well differentiated epistemological groups: *anticipated*, *ongoing* and *achieved* impact. Anticipated impact mostly refers to the description of the intentions that a specific educational tool has, and its resuls are of relevance for policy makers. On the other hand, ongoing impact refers to the evaluation of practices and is typically shared with staff development. Finally, achieved impact promotes particular kinds of outcomes and is relevant to support staff. Our research focuses on *ongoing impact* as we evaluate how impact

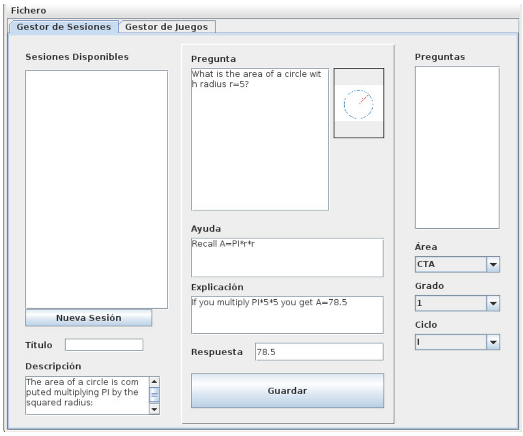

Figure 1. PC tool with a lesson created by a teacher.

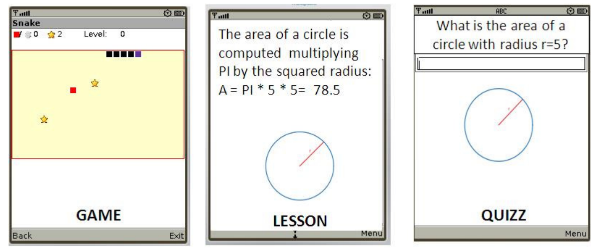

Figure 2. (a) A student eats an item in the Snake game (b) A lesson is shown (c) Quiz asking about content in the lesson

took place and the outcomes will be relevant to teachers willing to deploy a mobile learning program at their schools.

In terms of impact measurement, Sharples *et al.* distinguished between impact measured as learning gains of the student (*effectiveness*); impact measured as motivational changes (*satisfaction*, and impact measured as *usability of the tool* [17, 13]. For example, Vavoula *et al.* analyzed the educational impact of *Myartspace* through qualitative analyses based on structured diaries and interviews to measure both effectiveness and satisfaction in informal learning environments [23, 24]. On the other hand, Ramos *et al.* used a mixed-method approach with interviews and logs to evaluate the effectiveness of a mobile learning tool for non-formal, distance education [16]. Given that the usability of *EducaMovil* was already analyzed in [6], our research focuses on the use of quantitative analyses to evaluate educational impact in terms of effectiveness, both in formal and informal environments. Additionally, we also propose a qualitative evaluation through personal interviews with teachers and students so as to draw some hints and insights on the satisfaction of the tool.

LEARNING ENVIRONMENT

In this Section, we present a description of *EducaMovil* and we give an overview of the school and the setup of the pilot for the evaluation.

EducaMovil

EducaMovil is a platform that has two main components: (1) a PC tool for teachers to create educational contents and (2) a mobile game-based educational application for Java-enabled cell phones, which constitute more than half of the cells present in Latin American economies [5]. On the PC, teachers can create the educational contents that will be shown in the games [11]. On the cell phone, the mobile game-based application consists of an open-source cell phone game where points and lives are won after correctly answering a specific educational content. The PC tool allows teachers to bundle together educational content and game in a seamless manner: the teacher simply needs to select the set of lessons that wants to deliver with the game, and click on a button. The package is then automatically downloaded to all student cell phones via Bluetooth. Next, we explain the main components of the EducaMovil in detail.

PC Tool

The PC tool allows teachers to create the educational contents that will be shown in the mobile game. Each educational content is composed of a lesson and a quiz, although only quizzes are also possible. The lesson typically contains an image and/or an explanation about a specific concept. A lesson is always followed by a quiz, which can be a simple- or multiple-choice test question. Students' answers are used to evaluate their learning gains. For that purpose, their answers are compared against the correct quiz answers given by the teachers while creating the lessons. Additionally, the PC tool allows teachers to provide a *hint* that is shown whenever students press the *help* button during a quiz. Finally, teachers can also add an *explanation* that is displayed to the student once he runs out of chances to answer the quiz (each quiz is shown a maximum of three times). Figure 1 shows an example of the use of the PC tool by a teacher. As can be observed, the teacher has edited a lesson about how to compute the area of a circle, providing its answer, some hints in case the student asks for help, and an explanation on how to reach the correct answer, given the student fails to do so.

Figure 2 shows how the educational content created by the teacher is shown in the game after the student has eaten an item in the Snake game (2(a)). The lesson contains the explanation about the area of a circle written by the teacher (2(b)). Once the student feels confident with the concept explained, he/she can jump to the quiz (2(c)).

Cell phone Tool

The mobile phone application that students run on their cell phones consists of two elements: the game and the educational contents created by the teachers. EducaMovil offers the possibility of embedding the educational contents into different open-source games for cell phones like Snake, Tetris or Rally. The games are modified to add a module that handles the management of the educational contents. This module consists of three components: Game Model, Adaptation Model and User Model.

The *game model* is responsible for the interaction between the open-source game and the educational contents created by the teachers. The game model is fired every time an event in the game allows players to win points and introduces the educational contents as the units that need to be solved before winning the points. After the student finishes exploring the lesson, s/he will be prompted with a quiz, which is based on the lesson's content. If the learner provides a correct answer to the quiz, she is awarded one point. At any time, students can check their points and compare them against the top 10 highest scores.

The *adaptation model* determines the specific educational content that is going to be shown to the student at each step of the game. The educational contents are randomly selected from the pool of educational contents created by the teachers which are embedded into the game and downloaded to the cell phone. Each lesson and quiz are posed to the student up to three times, but never in a consecutive manner. If the student needs help, she can press the *help* button and the game will show a hint to help her answer the quiz. If the student does not successfully answer the quiz after three attempts, an explanation with the correct answer is shown and the lesson is eliminated from the pool of educational contents.

Finally, the *user model* stores the interactions of the student with the educational contents and the game. The model keeps counters about the lessons explored by the student, whether the quizzes were answered correctly or not, the time invested to answer quizzes, and whether the *help button* or the final *explanation* were instantiated. This model, which can be transfered to the PC tool via bluetooth, is primarily used for the quantitative analysis of the educational impact of *EducaMovil*, as shown in the Evaluation Section.

School and Pilot Setup

We have worked with a public school in a low-income peri-urban area of Lima, Peru (see Figure 3). The school, which offers both primary and secondary education, is part of the *Peru Educa* program run by the Peruvian Ministry of Education which focuses on providing educational tools and contents to teachers and students through an online portal [14]. For that purpose, the school has a very modest lab with second-hand computers where teachers and students can access contents online. The teachers told us that most students are highly motivated when classes are held in the *Innovation Lab*. However, given the large number of students at the school, the computer lab can be used only once a month by each class. Additionally, computers are often times infected by viruses and students have to share the PCs at a ratio of about five students per computer.

These issues pushed us to look for alternatives to the PCs that would be both engaging and attractive to the students while being more affordable and sustainable (in terms of maintenance). Given the high penetration rates of cell phone ownership at the school, we decided to explore the deployment of a mobile learning program. In order to understand the type of mobile learning tool that would better suit the curricular needs of the school, we hold interviews with both teachers and students. The teachers commented that they often test students with multiple-choice exams due to the characteristics of the Peruvian educational system. In Peru, students willing to move from secondary school to a public or private university are required to pass a multiple-choice exam that covers all curricular areas. Thus, teachers put a lot of emphasis on multiple-choice tests, specially in the secondary school, to prepare students for these highly competitive exams. Additionally, interviews with the students revealed that they were also very interested in having access to tests on their cell phones to be able to practice as much as possible; and more than 90% of the students declared that they used their cell phones mostly to call, send SMSs and to play games. These findings made us believe that *EducaMovil* could be the most adequate tool since it provides game-based learning on mobile platforms and the possibility to test students with multiple-choice quizzes.

The academic year at the school is divided into three terms and classes are run throughout the year covering different curricular objectives each term. At the end of each term, students take a final test -heavily based on multiple-choice questions- to evaluate their performance. We deployed our research pilot on the last term, from September'11 to December'11. The teacher for the *Science and Technology* class volunteered to run the pilot with two first grade groups at the secondary level ($ST1$ and $ST2$). During the last two weeks of August'11, we offered a workshop for the teacher to become acquainted with the PC tool and sessions with the students to learn how to use the mobile games. Although *EducaMovil* works on any Java-enabled feature phone, all students were provided with the same feature phone during the sessions to be able to focus our study exclusively on learning gains. Both the teacher and the students had help from a lab technician throughout the term. At the beginning of the pilot, we assigned each first grade group to one usage scenario: students in $ST1$ used *EducaMovil* in the classroom while students in $ST2$ used it during the school breaks. Students that participated in the classroom sessions, did not participate in the break environment and viceversa. The classroom setting constitutes a formal scenario where at the end of the class, and twice a week, each student was given a cell phone and was required to play during 20 minutes. The game always contains 10 lessons that the teacher prepares specifically for the content that has been presented during the class. Students were allowed to ask the teacher for help during the game sessions. On the other hand, the school breaks represent an informal environment that allowed students to go to the Innovation Lab twice a week during the breaks and play with *EducaMovil*. It is important to highlight that in this scenario students were not forced to play and it was up to them to decide when to go and for how long they wanted to play (the duration of the break was 30 minutes). However, the quizzes shown to them in every session were the same that the students in the classroom were shown. We believe that the school break setting can be used as a proxy to other informal scenarios like, for example, playing the game on a bus or while staying at home, which also are noisy environments and have no teacher's supervision. Finally, for control purposes, we also worked with two other first grade groups $C1$ and $C2$ that were taking the same class as $ST1$ and $ST2$. Students in $C1$ used the last 20 minutes of their class to solve paper-based quizzes at their own pace whereas students in $C2$ could voluntarily ask for paper-based quizzes during the breaks. These control groups were selected in such a way that the initial average grade was similar across all classes *i.e.,* the four groups start the pilot with a similar level of knowledge.

Figure 3. Classroom at the Principe de Asturias School, Lima, Peru.

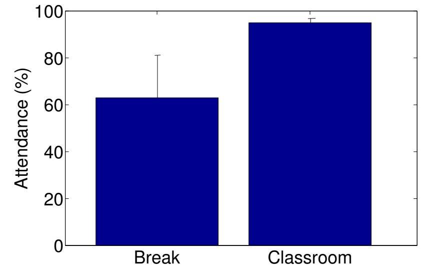

Figure 4. Average percentage of students that attend EducaMovil sessions in the classroom and during the breaks.

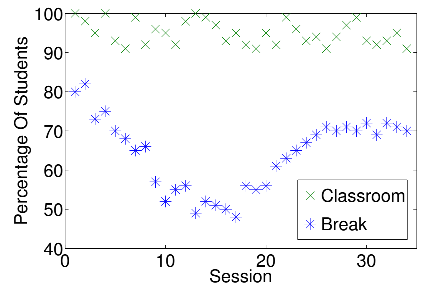

Figure 5. Student participation per session in the classroom (around 90%) and during the breaks (from 50% to 80%).

RESEARCH QUESTIONS AND METHODOLOGY

The focus of our research is to evaluate the educational impact of *EducaMovil* in terms of learning gains (*effectiveness*) when used in the classroom or during the school breaks. For that purpose, we carry out quantitative and qualitative analyses to reveal statistically significant learning outcomes and to gain insights into why these are taking place. Our end objective is to draw a set of guidelines that will help the teachers' community understand the educational impact of *EducaMovil* and decide about the convenience of deploying such tool at their schools. We formalize our study into three research questions.

The first research question (Q1) focuses on *analyzing whether the use of EducaMovil has an impact on knowledge acquisition, and if so, under which learning settings*. The fundamental underlying questions are whether students using *EducaMovil* can improve their level of knowledge more than students using paper-based tests; and whether informal environments like the school breaks promote more or less these gains than the classroom setting. We measure the level of knowledge with the final test grades that students get at the end of the term. In order to evaluate the educational impact, we collect the final test grades of the *Science and Technology* class for all the students using *EducaMovil* (groups $ST1$ and $ST2$) as well as for all the students in the control groups $C1$ and $C2$. Statistically significant differences in test grades between control and EducaMovil groups will determine that the tool probably has an educational impact on knowledge acquisition. Additionally, by measuring the difference between the initial and the final test grades in the classroom and break scenarios, we will be able to evaluate what type of environment favours more knowledge acquisition.

Our second research question (Q2) *evaluates what type of students benefit the most from using EducaMovil*. In fact, it is highly important for teachers to understand whether a minimum initial level of knowledge is necessary to experience significant individual learning gains or, on the contrary, whether even students with very low performance records can benefit from *EducaMovil*. We measure the prior level of knowledge of a student with his initial test grade before using *EducaMovil*, and the individual learning gain is computed as the percentage change between the initial and the final test grades of the *Science and Technology* class. In the end, we seek to understand if there exists an initial performance threshold for students in $ST1$ and $ST2$ to guarantee a significant learning gain in the classroom or the school break setting.

The third research question we pose (Q3) *analyzes the relationship between game performance and classroom performance so as to understand whether* EducaMovil *can be used as a proxy to measure a student's level of knowledge.* The final objective of this research question is to propose the use of EducaMovil as a paperless complement to student homework in order to help teachers with their evaluations. For that purpose, we study the relationship between the final test grades obtained by the students at the end of the term and their average game points across all game sessions. The existence or absence of strong correlations between the two distributions will reveal the potential of EducaMovil to approximate students' test grades. Finally, we use the outcomes from these research questions to build a framework for the semi-structured interviews that we carry out at the end of the pilot. By talking with the teacher and the students that used *EducaMovil*, we expect to gain a better understanding of the reasons behind some of our findings.

EVALUATION OUTCOMES AND INSIGHTS

This section first presents the datasets containing the student information we use for our analyses. Next, we describe general statistics about the usage of the tool and we close presenting our analytical results and insights for each research question.

Data Collection

Four groups participate in the evaluation: classes $ST1$ and $ST2$ use *EducaMovil* in the classroom and during the school breaks, respectively; and groups $C1$ and $C2$ act as control groups. Classes $ST1$ and $ST2$ are composed of 30 and 29 students from the first grade in secondary school; and classes $CT1$ and $C2$ have 28 and 32 students also from the first

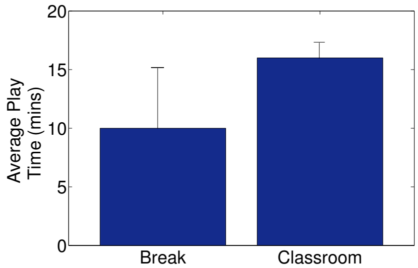

Figure 6. Average playing time with EducaMovil in classroom and break sessions.

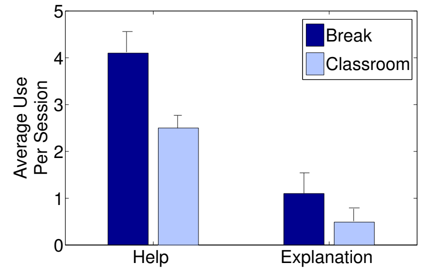

Figure 7. Usage of the *Help* button and instantiation of final Explanations.

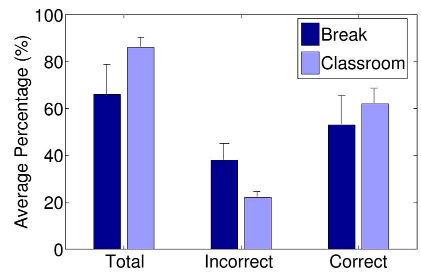

Figure 8. Average percentage of total quizzes answered, incorrect and correct ones both in the classroom sessions and during the breaks.

grade. Recall that all student groups were selected in such a way that the initial average grade was similar across groups, which allows us to compare them in a fair manner. For analytical purposes, we collect the initial and final *Science and Technology* test grades for each individual student in the pilot and control groups. The initial grades for the third term are the final grades obtained by each student at the end of the second term for the same subject.

EducaMovil was used by the students in $ST1$ and $ST2$ during a total of 34 sessions, which happened twice per week. For each session, the teacher prepared a set of 10 lessons that query about the curricular content presented by the teacher during the class. At the end of each session, we collected a user model for each student containing the number of lessons explored, the number of correct answers, time of exploration per lesson, and number of times the *help button* or the final *explanation* were instantiated. It is important to highlight that we did not have access to the results of the paper-based quizzes from $C1$ and $C2$ which they took home together with the results sheet. Although we missed analytical insights, we believe that this replicates best the anonymity that students feel while playing with *EducaMovil*. Additionally, at the end of the term, we also collected the notes from all the semi-structured interviews that we carried out with the teacher, the lab technician and the students.

General Statistics

Figures 4 and 5 reflect the student participation in both formal and informal learning environments. Figure 4 shows the average percentage of students that attend the *EducaMovil* sessions in the classroom and during the breaks. We observe that, on average, 95% ($\sigma = 1.2$) of the students in the classroom attended the game sessions whereas only an average of 63% ($\sigma = 23.2$) of the students played during the breaks. These statistics clearly show the differences in nature of both scenarios: in the classroom environment teachers require students to play during the last 20 minutes of the class which makes it almost impossible to miss the session, unless the student has not attended the class. On the other hand, students that use *EducaMovil* during the breaks, are encouraged but not forced to play. Figure 5 shows the participation of the students per individual session. We observe a high attendance of students during the first break sessions. However, it decreases over time until stabilizing at around an average of 65% of participation. Classroom students show consistently high participation rates throughout the pilot.

Figures 6, 7 and 8 delve into the game and quiz outcomes. Figure 6 shows that the average playing time of students in a classroom session is of $16.2m$ ($\sigma = 1.1m$) compared to the average playing time during the breaks which is shorter ($10.1m$ with $\sigma = 5.5m$). On the other hand, Figure 7 shows the average number of times per session students required help from EducaMovil as well as the average number of times the tool output an explanation. We observe that students in the classroom use less the help button (2.5 times per session on average) and reach less times the final explanation (0.49 times per session), probably due to the fact that they can ask their teacher who is present while the session takes place. On the other hand, students that play during the breaks use more the help button (an average of 4.1 times per session) and reach more times the final explanation (1.1 times on average per session). Since the final explanation is shown only when the student has given an incorrect answer three times for the same quiz, these findings reveal that students during the breaks have more problems reaching final correct answers.

Figure 8 shows that students playing in the classroom give more answers, correct or incorrect, than students playing during the breaks. On average, classroom students answer an 86% ($sigma = 6\%$) of the 10 quizzes that the game session contains as opposed to the 66% ($\sigma = 16\%$) of the quizzes answered by students during the breaks. Additionally, students that play during the breaks make a higher percentage of mistakes in their answers than students playing in the classroom, precisely a 16% more (38% versus 22%). Recall that these are just incorrect answers meaning that the student might answer a quiz wrong the first time but correct the next one. In

Group Final Grade	Significance
ST1 vs. C1,C2	$p = 0.04$ $(t = 2.14, df = 56; t = 2.13, df = 60)$
ST1 vs. ST2	$p = 0.03$ $(t = 2.26, df = 57)$
ST2 vs. C1,C2	-

Table 1. Statistical difference in grades between control groups and students that participated in the pilot.

terms of percentage of correct answers, we see that on average students in the classroom answer correctly a 62% ($\sigma = 12\%$) of the questions whereas during the breaks the percentage of correct answers goes down to 53% ($\sigma = 16\%$). These general statistics show that, on average, students in the classroom seem to use *EducaMovil* more frequently and with better results than students playing during the breaks. Next, we answer the research questions that will allow us to delve more into these observations.

Q1: Educational Impact

To evaluate the impact of *EducaMovil* in terms of *effectiveness*, we run statistical *t-tests* between the final test grades of the control groups and the groups that used *EducaMovil* both in the classroom or during the breaks. The underlying assumption is that improvements in the final test grades are related to the use of *EducaMovil* as opposed to paper-based quizzes. This is a reasonable assumption given that a game-based approach might engage students more through interactivity and practice repetition thus increasing learning gains. Since the student groups are sufficiently large, and since they represent independent sample groups, we use *unpaired t-tests*. Recall that all groups start with a similar level of knowledge *i.e., t-tests* between the initial test grades across all groups did not reveal statistical differences. As shown in Table 1, we observe statistically significant differences between the final test grades of the classroom $ST1$ and the control groups, as well as with the break group $ST2$, at a p-value $p < 1\%$. However, we do not observe any statistical difference between the final test grades of the control groups and the group of students that played with *EducaMovil* during the breaks ($ST2$). This analysis shows that *EducaMovil* might enhance knowledge acquisition when used in the formal environment of a classroom. In order to quantify its impact we applied a *paired t-test* between the students' test grades at the beginning (pre-use) and at the end of the term (post-use). The test showed statistically significant differences with $p < 1\%$. The average pre-use grades were of 10.2 and the post-use grades increased to 11.6 (in a scale of $0 - 20$ with passing grades starting in $7 - 10$).

Next, we delve into the possible reasons behind the lack of impact of the tool when used during the school breaks. There are two elements that differentiate the break environment from the classroom environment: (i) not all students attend all sessions *i.e.,* some students might go very often while others might go from time to time unlike the classroom environment where all students are required to attend the sessions; and (ii) not all students play the same amount of time. In the classroom, all students play for $20 minutes$ which is the time allocated by the teacher, whereas during the $30 - minute$ breaks students are free to play as much time as the want.

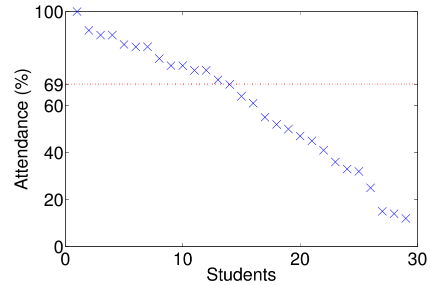

Figure 9. Percentage of sessions that each student attended throughout the duration of the pilot during school breaks. With attendance above 69% knowledge acquisition becomes statistically significant.

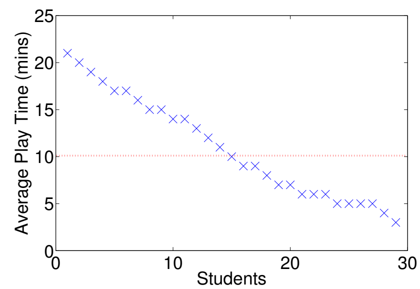

Figure 10. Average playing time per student during school breaks. Students that play for at least 10 minutes per session show statistically significant learning gains when compared to the control groups.

Figure 9 shows the percentage of sessions that each student in $ST2$ attended during the breaks (sorted in descending order). We observe that some students attended almost all sessions whereas others attended very few. Thus, it is critical to evaluate whether there exists a minimum number of sessions that students need to attend in order to show significant improvements in knowledge acquisition. For that purpose, we build distributions of students that use *EducaMovil* for at least a minimum percentage of sessions x (with $x \in (0, 100)$) and compare them to the control groups. In this case, given that the sizes of the distributions might be really small and of different sizes, we run the non-parametric version of the *t-test*, the *Man-Whitney test*. The tests highlight that there exist statistically significant differences between the final test grades of the control groups and the students using *EducaMovil* during the breaks when these attend at least 69% of the sessions (with the mean grade of the control groups being smaller). This analysis suggests that *EducaMovil* might also have an educational impact when deployed in informal environments as long as students use it for a minimum number of sessions.

Figure 10 shows the average playing times per session for the students that used *EducaMovil* during the breaks (sorted in descending order). We observe that there are students that use the tool for long intervals whereas others stop playing after a few minutes. In order to understand the role of the average playing time on the educational impact of the tool, we first build distributions of students with average playing times of at least x minutes ($x \in (0, 30)$). Next, we run *Man-Whitney*

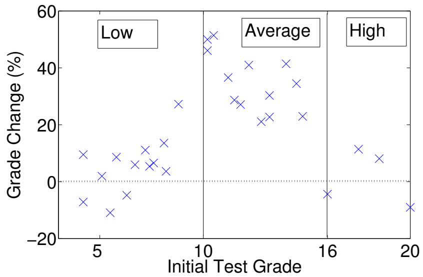

Figure 11. Percentage of change between the initial and the final grades for each group of students that use *EducaMovil* in the classroom.

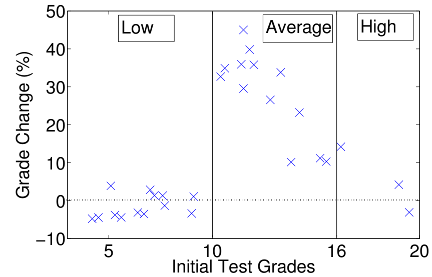

Figure 12. Percentage of change between the initial and the final grades for each group of students that use *EducaMovil* during school breaks.

tests between the final test grades of the students in these distributions and in the control groups and evaluate the differences. The tests showed statistically significant differences (with lower mean grades for the control groups) when the students use *EducaMovil* for at least 10 minutes per session on average (which corresponds to exploring at least 6 quizzes). To conclude, we propose **Guideline #1**: *EducaMovil* can improve knowledge acquisition in both formal and informal environments. However, in informal environments, it is important to make sure that students interact with the tool for a minimum amount of time, otherwise the educational impact will not be statistically significant.

Q2: Who benefits the most?

To analyze what type of students benefit the most from using *EducaMovil*, we divide the students in $ST1$ and $ST2$ into three groups based on their initial test grades before using the tool: high-grade group (from 16 to 20), average-grade group (from 10 to 16) and low-grade group (from 1 to 10). We focus our analysis on evaluating which group of students experiences the largest learning gains. As with $Q1$, the underlying assumption is that improvements in the test grades are related to the use of *EducaMovil* and imply knowledge acquisition. Specifically, we represent individual learning gains as the percentage of change between initial g_0 and final grades g_f *i.e.,* $100*(g_f - g_0)/g_0$ and use this value as a proxy representation of the benefits of the tool.

Figure 11 shows the percentage of change between the initial and the final grades for the students that use *EducaMovil* in the classroom. In general, we observe that the majority of the students show a positive change in their grades *i.e.,* their final test grades are typically higher than their initial test grades. The negative grade changes are of at most -10% which means that the students that obtained worse grades at the end of the term, experienced a decrease of at most two points in their final grades. In terms of groups, we observe that students with initial high grades do not experience large changes in their final grades. This is probably due to the fact that their grades are already high and there is very little room for improvement. At the other extreme, students with initial low grades experience improvements in their final grades of at most a 13% (with the exception of one student that shows an increase of a 22%). Such improvement translates into improvements in the final test grades of approximately two points after using *EducaMovil*. Finally, we observe that the students that show the highest percentages of improvement in their final grades are the students with average initial grades. In fact, this group has students with improvements from 20% all the way up to 50%, which means that the final test grades have incremented at most five points.

On the other hand, Figure 12 shows the percentage of change between initial and final grades for the students that used EducaMovil during the school breaks. We observe that, similarly to the classroom environment, students with high initial grades do not show an important change in their grade evolution over the duration of the pilot. Students with low initial grades, show grade changes of up to 5% which translates into increments of at most one point in the final test grades. Compared to the low-grade group in the classroom environment (see Figure 11), the percentage of change in the school break setting is considerably smaller (5% versus 13%). Digging more into these low-grade students, we observed that they all shared lower average playing times and lower session attendance than their classroom counterpart, confirming the need to require a minimum participation so as to guarantee an educational impact (Guideline #1). Finally, students with average initial grades are again the ones that appear to benefit the most from using *EducaMovil*, showing percentages of grade change of up to 45%. In general, the learning gains for each student group (low, average and high) in the break scenario are smaller than the gains for students using *EducaMovil* in the classroom. Unfortunately, we cannot report on the statistical significance of this observation given the small size of each group. However, differences in percentage of learning gains between the classroom and the breaks are around 5%. To sum up, we propose **Guideline #2** for teachers: students that start using *EducaMovil* with an initial average grade are the ones that benefit the most from using the tool. Thus, before deploying *EducaMovil*, it is important to check that the students have a minimum working knowledge base of the subject to guarantee a significant educational impact. Additionally, teachers should bear in mind that classroom deployments appear to produce higher learning gains –across all groups– than more informal environments.

Q3: EducaMovil and Classroom Performance

To evaluate whether *EducaMovil* can be used as a proxy to measure a student's level of knowledge, we analyze

the relationship between test performance and game performance across all the students using the tool. Test performance

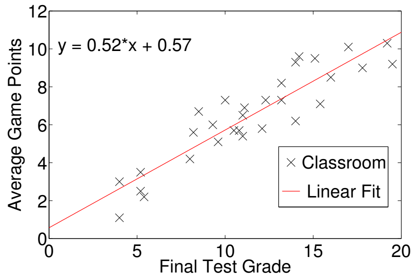

Figure 13. Linear regression of test grades and average game points for students that used *EducaMovil* in the classroom. $R^2_{adj} = 0.86$**.**

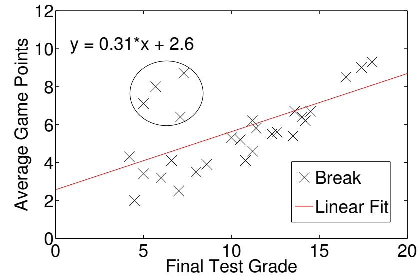

Figure 14. Linear regression of test grades and average game points for students that used *EducaMovil* during school breaks. $R^2_{adj} = 0.64$**.**

is represented as the final test grade and game performance with the average game points obtained by the student throughout all *EducaMovil* sessions. We study the relationship between the two variables applying linear regression (with the least squares approach) and Pearson's correlation analyses.

Figure 13 represents the final test grade and the average game points for each student using *EducaMovil* in the classroom, and its linear regression. We observe that the trendline appears to fit the data quite well, with a residual sum of squares of $R^2_{adj} = 0.86$. The correlation analysis between the final grades and the game points distributions shows a strong relationship with a coefficient of $r = 0.87$. The linear relationship as well as the strong correlation determines that students with high test grades (17 and above) show high average points per game (around 7 on average); students with low test grades (8 and below) share low average game points (4 and below); and students with average test grades (between 8 and 16) have average game points between 4 and 7.

On the other hand, Figure 14 shows the relationship between final test grades and average game points for the students that used *EducaMovil* during the breaks. We observe that the linear fit is not as good as the classroom trendline in Figure 13, with a residual sum of squares (adjusted R^2) of 0.64. The correlation analysis reveals a medium correlation coefficient of $r = 0.6$. These results are probably due to the fact that there exists a group of students (grouped within a circle in the Figure), that have low test grades but show high average game points. We analyzed each of the students in this group and observed that all four rarely went to any of the sessions (less than 20% of the times). Their high average game points are the result of having played only two or three sessions and having provided a high percentage of correct answers. However, given the small number of sessions they participated in, it is hard to determine whether the high game points are due to randomness or to the students' knowledge. In order to eliminate the impact of these spurious outliers on the fitting, we suggest adding a filter that requires a minimum percentage of session attendance per student. By doing so, the linear fitting in our analysis highly improves with a residual sum of squares $R^2_{adj} = 0.79$ and a correlation coefficient of $r = 0.81$. We suggest **Guideline #3**: *EducaMovil* game points can be used as a proxy for student test grades in the classroom environment; in the break environment, it acts as a proxy only when attendance requirements are put into place. Thus, teachers could potentially use *EducaMovil* as a complementary tool to homework so as to help them with student evaluation.

Qualitative Analysis

To gain insight into some of the previous findings, we designed semi-structured interviews based on the main outcomes from the quantitative analysis. At the end of the pilot, we conducted individual interviews with the teacher, the lab technician and a total of 20 students with different levels of knowledge. Next, we describe the interview questions and highlights from our conversations.

Research questions *Q1* and *Q3* have revealed that the educational impact of *EducaMovil* in terms of learning gains highly depends on a minimum engagement of the student with the tool. In fact, students that used *EducaMovil* frequently and during long periods of time appear to benefit more than students that had a scarce interaction. Although the classroom environment *forces* students to participate in two sessions per week, during the breaks students are free to decide whether to play or not. In an attempt to guarantee a minimum engagement also for students that use *EducaMovil* during the school breaks, we seek to understand what motivated some to engage frequently and what prevented others from doing so. In our interviews, the teacher and the lab technician explained that most of the students participated during the first break sessions of *EducaMovil*. However, they observed that for some students the motivation fell as time passed. They hypothesize that some of the students that failed to participate in the long run, might have felt *left behind* and discouraged because they were not doing as well as their peers *i.e.,* in a competitive game, they were getting less points. The teacher suggested that these students might have felt ashamed and thus might have stopped attending the sessions. They recommended that it would be advisable to personalize the quizzes and adapt them to each student's level of knowledge so as to maintain the engagement throughout the semester. On the other hand, students with low participation records mentioned in their interviews that the questions in the game were too difficult, which might confirm the hypothesis that they felt discouraged because they could not answer many of the quizzes. Although some students in the classroom environment shared the same feeling, these tended to ask their teacher to help them with the quizzes thus maintaining their motivation.

Research question *Q2* identified that students with initial low grades do not benefit as much from *EducaMovil* as students with average initial grades, consistently showing better results in the classroom setting. To understand possible reasons behind this, we interviewed low-grade students who complained about the complexity of the quizzes, which sometimes went beyond class content and required a little bit of research on the students' side. The teacher, in fact, did put some quizzes that required the students to go through the book, find information and reason a little bit before being able to provide an answer. This thinking process might have put back students with lower levels of knowledge for whom researching for an answer could be more difficult. However, low-grade students in the classroom could ask their teacher and engage in conversations whereas students that played *EducaMovil* during the breaks did not have access to teachers or books and played in noisier environments, which might have also affected their concentration. To overcome this problem, we observed that during the breaks low-grade students used much more the help button than the students in the classroom environment. Unfortunately, the help that *EducaMovil* gives, focuses on providing the solution instead of explaining how to do the research to find a specific answer. Thus, the teacher suggested that the help button could be enhanced by engaging the student in an *informal* conversation with an avatar or a character in the game to provide supervision and guidance. These insights will be integrated into future versions of *EducaMovil* which will incorporate personalization techniques to adapt the complexity of the quizzes to each student's knowledge base. Such adaptation will hopefully solve the motivation problem that some students had during the breaks where no teacher supervision was provided. Additionally, we will investigate ways to improve the help offered in *EducaMovil* by stimulating the student towards an inquiry process with a character in the game, instead of simply providing the formula for the answer.

CONCLUSIONS

The pervasiveness of feature phones in emerging economies has contributed to the advent of mobile learning applications for low-income populations. Our research extends the state of the art by presenting a quantitative and qualitative evaluation of a game-based mobile learning tool deployed in both formal and informal learning settings at a low-income school in Lima. We have provided teachers with a set of guidelines for the successful implementation of *EducaMovil*. Among our main findings, we have shown that EducaMovil has an educational impact both in formal and informal environments as long as a minimum interaction with the tool is guaranteed. We have also shown that students with initial average grades benefit the most from using the tool and we have demonstrated that *EducaMovil* could be used as a paperless complement to homework for student evaluation. We finish proposing a set of improvements for *EducaMovil* based on our qualitative assessments.

ACKNOWLEDGEMENTS

We would like to thank Justino Rivero, Sumbi and Telefonica Foundation for all their support.

REFERENCES

1. Alismail, H., and et al. Combining web technology and mobile phones to enhance english literacy in underserved communities. In *First Annual Symposium on Computing For Development* (2010).
2. BBC, Banglalink, Citycell, Grameenphone, TeleTalk, Robi, and Warid. **http://www.bbcjanala.com/**.
3. Butgereit, L. Math on mxit: using mxit as a medium for mathematics education. In *Meraka INNOVATE Conference for Educators* (2007).
4. Corlett, D., Sharples, M., Chan, T., and S., B. Evaluation of a mobile learning organiser for university students. *Journal of Computer Assisted Learning 21* (2005), 162–170.
5. Frias, V., J.Virseda, and Frias, E. The anatomy of mobile handsets: On the development of effective cell phone services. In *Mobile Communication Technology for Development (M4D)* (2010).
6. Frias-Martinez, V., and Virseda, J. Personalization of mobile learning tools for low-income populations. In *Personalization Approaches in Learning Environments Workshop, PALE'11 (UMAP)* (2011).
7. Kalloo, V., and Kinshuk, M. Personalized game based mobile learning to assist high school students with mathematics. In *IEEE International Conference on Advanced Learning Technologies* (2010).
8. Kam, M., and et al. Improving literacy in rural india: Cellphone games in an after-school program. In *ACM Conference on Information and Communication Technology and Development* (2009), 139–149.
9. Ketamo. An adaptive geometry game for handheld devices. *Educational Technology & Society 6(1)* (2003).
10. Malone, T. What makes things fun to learn? In *3rd ACM SIGSMALL* (1980), 162–169.
11. Molnar, A., and Frias-Martinez, V. Educamovil: Mobile educational games made easy. In *Proceedings of the World Conference on Educational Multimedia, Hypermedia and Telecommunications* (2011).
12. Ogawa, M. Augmenting large-enrollment instruction with mobile learning. In *Annual Technology, Colleges and Community* (2010).
13. Oliver, M., and Harvey, J. What does impact mean in the evaluation of learning technology? *Educational Technology and Society 5(3)* (2002).
14. PeruEduca. **http://www.perueduca.edu.pe**.
15. Price, S., and Oliver, M. A framework for conceptualising the impact of technology on teaching and learning. *Educational Technology and Society 10(1)* (2007), 16–27.
16. Ramos, A. J. Project mind: The viability of mobile sms technologies for non-formal distance learning in asia. *Philippines: Molave Development Foundation, Inc.* (2008).
17. Sharples, M. Methods for evaluating mobile learning. *Researching Mobile Learning: Frameworks, Tools and Research Designs. Oxford: Peter Lang Publishing Group* (2009).
18. Tian, F., and et al. Let's play chinese characters: Mobile learning approaches via culturally inspired group games. In *International Conference on Human Factors in Computing Systems* (2010).
19. Toyama, K. Can technology end poverty? *Boston Review 36(5)* (2010).
20. Traxler, J., and Kukulska-Hulme, A. Evaluating mobile learning: Reflections on current practice. In *mLearn 2005. Mobile technology: The future of learning in your hands* (2005).
21. Trucano, M. Worst practice in ict use in education, 2010. EduTech, World Bank Blog.
22. Trucano, M. Mobile learning in developing countries in 2011: What's new, what's next?, 2011. EduTech, World Bank Blog.
23. Vavoula, G., and et al. Myartspace: Design and evaluation of support for learning with multimedia phones between classrooms and museums. *Computers and Education 53*, 2 (2009).
24. Vavoula, G. N. A study of mobile learning practices, report of mobilearn project. **http://www.mobilearn.org/download/**.
25. W., D. Building mathematical knowledge in an authentic mobile phone environment. *Australasian Journal of Educational Technology 26*, 1 (2010), 85–104.
26. Yin, C., Ogata, H., and Yano, Y. Japelas2: Japanese polite expressions learning assisting system in ubiquitous environments. In *Proceedings of Supporting Learning Flow through Integrative Technologies* (2007).