Personalization Of Mobile Learning Tools For Low-Income Populations

Abstract. The ubiquitous presence of cell phones in emerging economies has converted them in ideal platforms to cater services for underserved communities in areas like mobile learning. M-learning tools have proved successful in such challenging environments, specially for afterschool and vocational programs. A key component of that success is the personalization of the tools to the community and to particular individual needs. However, in order to tailor contents to the students, we first need a deep understanding of their learning abilities and preferences. In this paper, we propose design guidelines for the personalization of a mobile learning tool for low-income populations called EducaMovil.

1 Introduction

There exists a large collection of mobile learning tools for low-income communities adapted to all types of cell phones, mostly deploying SMS- or Java-based solutions to deliver educational content [1]. Our research focuses on the personalization of such educational tools, which is specially critical in emerging economies where access to education tends to be more limited and irregular. As a result, students with similar ages might show very different levels of achievement and could thus benefit from personalized contents. The delivery of personalized education is complex because the adaptation to each individuals' requirements demands a deep understanding of their abilities and preferences. Related work in the areas of e-learning and tutoring systems has demonstrated that demographic factors and learning capabilities play an important role in knowledge acquisition [3]. Nevertheless, very little work has been done to understand which factors impact the learning process in the area of mobile learning for low-income communities [4]. In this paper, we present an evaluation of the repercussion that human demographics and individual learning abilities might have on the successful use of *EducaMovil* in a low-income community [5]. The results of this evaluation will provide design guidelines for the personalization of our mobile learning tool prior to a future long-term evaluation of its educational impact.

2 Learning Environment and Data Collection

EducaMovil is a system that has two main components: (1) a PC tool for educational content creation and (2) a mobile game-based educational application for Java-enabled cell phones [5]. On the PC, teachers can create the educational content snippets that will be shown in the games. Each educational snippet is composed of a lesson and a quiz, and is labeled by the teacher with its educational grade and its complexity level. On the cell phone, the mobile game-based application consists of three models: (1) the *game model* is fired every time an event in the game allows players to win points and introduces the educational snippets that need to be answered; (2) the *adaptation model* determines the specific educational content that is going to be shown to the student at each step of the game based on his/her previous interactions; and (3) the *user model* stores the interactions of the student with the educational units and the game.

The evaluation of EducaMovil consisted of one-on-one sessions where each student sat down and received an initial description of the game, its rules, the educational snippets and how to navigate through these with the cell phone. After the initial introduction, we let the student play for 20 minutes. During the session, students were shown educational snippets from their own educational grade starting with easy lessons which evolved in complexity based on the adaptation model. A total of 27 students from a low-resource school in Lima (Peru) tested EducaMovil: 5 from each 1st, 2nd and 3rd years and 6 students from each 4th and 5th years of the secondary school. These students had ages between 12 and 16 and in terms of gender, 15 were female and 12 male. At the end of each session, we collected a user model with the student interactions and computed a *general performance model (GPM)* for each student i as $GPM_i = (C, T) = (100 * (\sum_{j=0}^{n} A_j)/n, (\sum_{j=0}^{n} T_j)/n)$ where A_j is the answer given to question j, T_j the time used to answer it and n represents the total number of lessons explored by the student.

3 Analysis and Game Design Implications

We carry out two types of analysis: (i) analyze whether there exist differences in the learning performance of the students based on gender or age and, (ii) evaluate the types of learning behaviors observed across all students. The first analysis will give us suggestions for gender- or age-based personalization. As for the second, it will provide design guidelines to personalize education based on *stereotypes i.e.,* learning behaviors shared by groups of students [2]. Although the results are preliminary and specific to our pilot, the techniques proposed can be used to guide the personalization of any mobile learning tool.

Gender and Age In order to understand whether gender or age impact the learning interactions of the students with the tool, we build gender- and age-based distributions for each performance variable (percentage of correct answers C and average time per lesson T) and compute statistical tests to understand whether the differences we observe are statistically significant or not. Given that our distributions are small (27 student models), we report results for non-parametric statistical methods to avoid assuming a normal distribution for a dataset that might not be sufficiently large. To carry out the **gender analysis**, we first compute the female and male distributions separately for each performance variable C and T. To test the differences between each pair of female-male distributions we use the *Kolmogorov-Smirnoff test* and, given the small number of samples we have, reject the *null hypothesis* whenever $p <= 0.1$. Our results show that the percentage of correct answers for females is higher than its male counterpart and statistically significantly different ($p = 0.08$). Specifically, the female distribution had an average percentage of correct answers of 58.4% ($\sigma = 18.3\%$) and the male had an average of 50.2% ($\sigma = 15.9\%$). On the other hand, we also observe that the female average answering time per quiz is also higher than its male counterpart and statistically significant ($p = 0.1$). Females showed an average answering time of $36.4s$ ($\sigma = 16.3s$) and males had an average time of $24.7s$ ($\sigma = 13.3s$). These numbers show that female students achieve statistically significant higher percentages of correct answers ($\approx 8\%$ more) but need more time to answer the quizzes ($\approx 12s$ more on average). This fact is not necessarily bad if it is related to women being more thoughtful, but could be harmful if connected to a lack of confidence. In an attempt to decrease answering times while maintaining performance, we suggest to put counters and timers in the quizzes so as to create a healthy competition between men and women not only in terms of correct answers but also in terms of answering times.

To perform the **age analysis** we compute for each performance variable C and T one distribution per educational level (age). To understand whether there exist differences in the student performance across educational levels (age groups), we run *Kruskal-Wallis* tests for each group of five distributions representing the five age groups present in the pilot. We reject the *null hypothesis* with $p <= 0.1$. Our results determined that there exists a statistical significant difference between the percentage of correct answers among some of the five educational levels with $p = 0.05$. However, we do not observe any statistically significant differences on the average answering time. We observe that students from the $3rd$ grade outperform all their peers with an statistically significantly different percentage of correct answers ($median = 85\%$), whereas students from the first grade show the worst performance with a median value of 29%. This analysis shows that although students are shown quizzes adapted to their own educational level, not all groups respond equally. In fact, we observe that students from educational levels $4th$ and $5th$ are statistically significantly outperformed by the students in the $2nd$ and $3rd$ grades. This might be related to learners in their last school years being allowed to move to the next educational level without making sure they have acquired the minimum required knowledge. We suggest to add quizzes from previous educational levels until the students show an improvement in their performance. This approach will allow them to review previous contents and to eventually reach their own educational level.

Stereotypes In this section, we use the *k-means* clustering technique to identify common learning behaviors (*stereotypes*), independent of age or gender, among the students in our study. Given that the final k-means partition highly depends on the initial seeds selected, we run the algorithm 100 times for each value of k, and select the cluster distribution with the most compact and well separated clusters *i.e.,* the one with the minimum *cluster validity index* computed as the

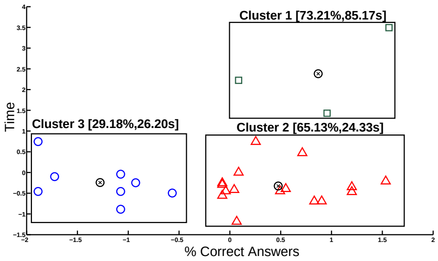

Fig. 1. Student stereotypes with $k = 3$ (axes represent normalized values).

ratio between the intra-cluster distance and the inter-cluster distance among all sample. Once the best selection of clusters has been identified for each value of k, we select the k with the smallest cluster validity index across all partitions evaluated (from $k = 2$ to $k = 7$). In practical terms, we attempt to find clusters of common performance behaviors across the 27 GPMs. For that purpose, we normalize across all GPMs, apply k-means and validate for each value of k. Although the minimum cluster validity index corresponds to $k = 2$, we discuss $k = 3$ since it provides more insight into the learning stereotypes than $k = 2$. Larger values of k have larger cluster validity indices and do not provide any other relevant information after exploration. Figure 1 shows the results for the clustering of the 27 GPMs with $k = 3$. *Cluster 1* with centroid $(73.2\%, 65.1s)$ and three students represents a group with the highest percentage of correct answers and the largest answering times. Interestingly enough, this group contains only female students. *Cluster 2* with centroid $(65.1\%, 24.3s)$ and 15 students, represents a group that employs little time to give the correct answers. This cluster probably groups the best students. *Cluster 3* with centroid $(29.1\%, 26.2s)$ represents a group of 9 students that share a low percentage of correct anwers and low average answering times. This group probably represents students who answer randomly with their interest devoted to the game instead of the educational contents. We suggest that when this type of behavior is identified on a specific student, the game should be slowed and show more than one quiz at a time to force the student to focus on the quizzes.

References

1. BBC. `http://www.bbcjanala.com/`.
2. P. Brusilovsky. Methods and techniques for adaptive hypermedia. *User Modeling and User-Adapted Interaction*, 6(2-3):87–129, 1996.
3. E. Frias-Martinez, S. Chen, and X. Liu. Investigation of behavior and perception of digital library users. *Journal of Information Management*, 28(5):355–365, 2008.
4. A. Kumar, A. Tewari, G. Shroff, M. Kam, and J. Canny. An exploratory study of unsupervised mobile learning in rural india. In *ACM CHI*, 2010.
5. A. Molnar and V. Frias-Martinez. Educamovil: Mobile educational games made easy. In *Proceedings of the World Conference Ed-Media*, 2011.