Computers & Education 55 (2010) 1069–1079

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

An experiment for improving students performance in secondary and tertiary education by means of m-learning auto-assessment

Luis de-Marcos*, José Ramón Hilera, Roberto Barchino, Lourdes Jiménez, José Javier Martínez, José Antonio Gutiérrez, José María Gutiérrez, Salvador Otón

Computer Science Department, University of Alcalá, Ctra Barcelona km 33.1, 28871 Alcalá de Henares, Madrid, Spain

A R T I C L E I N F O

Article history:
Received 29 September 2009
Received in revised form
7 May 2010
Accepted 7 May 2010

Keywords:
E-learning
Mobile learning (m-learning)
Mobile assessment

A B S T R A C T

Mobile learning is considered an evolution of e-learning that embraces the ubiquitous nature of current computational systems in order to improve teaching and learning. Within this context it is possible to develop mobile applications oriented to learning, but it is also important to assess to what extent such applications actually work. In this paper we present a new tool designed to reinforce students' knowledge by means of self-assessment. Improvement in student achievement was evaluated and an attitudinal survey was also carried out to measure student attitudes towards this new tool. Three different experimental groups were selected for this research, with students aged from 14 to 21 years old, including high-school and university students. Results show that this kind of tool improves student achievement, especially amongst younger learners, with a relatively low impact on current teaching activities and methodology.

1. Introduction

With the new "anytime, anywhere computing" paradigm (ubiquitous computing), a shift from "electronic" to "mobile" services has begun. Thus, just as e-commerce has extended to m-commerce, so e-learning now includes m-learning (mobile-learning) (Lehner & Nösekabel, 2002). In the field of teaching and learning, the expected benefits of this new mobility include, among others, more efficient instruction together with an improvement in learning outcomes. Within this framework it is crucial to create new tools that add value to the teaching-learning process, but it is also important to assess performance using such instruments in order to enable us to exert some control over the expected results of the learning process. In this context, this paper presents a new mobile application designed for self-assessment. It allows students to test their knowledge and expertise in a specific topic with questionnaires designed by their teachers. Young students use mobile phones as an integral part of their life, and consider them a crucial element in their communication and social activities (Funk, 2004). Therefore, providing them with learning tools that operate within this mobile environment is important since this will increase their motivation to learn and their engagement in learning activities (Attewell & Webster, 2004). However, designing and implementing new tools is not sufficient; such tools need to form an integral part of learning activities and their usefulness must also be measured. For these reasons, this paper also describes research carried out in order to test the new application in three different environments. Three groups of different ages and levels were chosen for experimentation and analysis. In Section 2, we present a brief review of state of the art of m-learning, together with the aims of this research. In Section 3, the mobile application and the system providing appropriate support to the learning action are presented. In Section 4, we describe the experiments that were carried out; and finally, Section 5 includes the results, analysis and conclusions.

2. Background & motivation

A number of definitions, covering a wide range of aspects, have been used to explain the term "m-learning". Some of them identify m-learning as simply an evolution of e-learning, while others define it as an independent trend that has its origins in the ubiquitous nature of

* Corresponding author. Tel.: +34 918856656; fax: +34 918856646.
E-mail addresses: luis.demarcos@uah.es (L. de-Marcos), jose.hilera@uah.es (J.R. Hilera), roberto.barchino@uah.es (R. Barchino), lou.jimenez@uah.es (L. Jiménez), josej.martinez@uah.es (J.J. Martínez), jantonio.gutierrez@uah.es (J.A. Gutiérrez), josem.gutierrez@uah.es (J.M. Gutiérrez), salvador.oton@uah.es (S. Otón).

0360-1315/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2010.05.003

present day communication systems. These identify m-learning as 'location-independent and situation-independent' (Nyíri, 2002). McLean (McLean, 2003) considered 'm-learning' to be a term coined to cover the wide range of applications which have emerged from the convergence of new mobile technologies, wireless infrastructure and e-learning developments. Further analysis of available definitions allows us to summarize the two essential features of m-learning: (1) being mobile, m-learning enables the educational process to take place anywhere and at anytime; and (2), any kind of handheld device (small and easy to carry) together with a communication technology, is required.

A recurrent theme in different studies on m-learning (Chen, 2010; Huang, lin, & cheng, 2010; Mobilearn, 2003; Motiwalla, 2007; Vavoula, Lefrere, O'malley, Sharples, & Taylor, 2004) is that mobile/electronic education should not attempt to replace traditional education, but should instead support both students and teachers by providing them with services that facilitate teaching, learning and/or any related administrative tasks. The basic approach is integrative, combining a variety of (mobile and non-mobile) devices and using either wired or wireless transmission technologies (Lehner & Nösekabel, 2002). This premise is also supported by Houser et al. (Houser, Thornton, & Kluge, 2002). After analysing successful m-learning projects, they concluded that all the projects they had studied used mobile devices as part of a mixed educational programme (b-learning or "blended learning") combining traditional attendance-based education with web learning and mobile components.

Shepherd (Shepherd, 2001) proposes three possible uses for m-learning: (1) the first is to use m-learning as an aid in the preparatory phase, before any learning actually takes place, through the use of "diagnosis". This includes pre-tests, learning-style tests, attitudinal surveys and the gathering of pre-requisite data concerning the learner's experience, job, and qualifications. This useful data may be used to avoid loss of time during teaching by adapting the learning experience to each learner profile. (2) The second use is to employ m-learning approaches as a means of supporting students when they are preparing for their examinations, reviewing contents and reinforcing the knowledge they have acquired to date. Finally (3), the most interesting challenge for m-learning, according to Shepherd, is the contribution it can make to continuous on-demand learning (usually applied to real-world problems). Another issue that must be considered is the kind of contents that can be delivered by means of m-learning. As Wuthrich et al. point out (Wuthrich, Halverson, Griffin, & Passos, 2003), the particular characteristics of the mobile devices used in this type of initiative mean that they can be used as a conduit for distributing self-evaluation tools and study guidelines, and in some cases enable feedback between educators and learners. These authors emphasize the essential role that tests and questions play in knowledge acquisition and suggest that mobile devices are especially suitable for questionnaire completion, given the increased mobility of students nowadays.

More recent field studies have focused on different aspects related to teaching and learning in a mobile environment, but offer little comparative analysis of the effect that similar systems have at different educational levels. For the purposes of organization we have classified this research into studies that present and analyze communication tools, studies that focus on content presentation and delivery, and studies centred on mobile assessment tools and methods. In the *first group*, Motiwalla (Motiwalla, 2007) presents a framework for mobile learning and a wapsite which provides different support tools including RSS news alerts, a discussion board, and a chat room. A pilot study was conducted with university students, and evaluation of their performance provided valuable information about the potential usefulness of such tools, but also revealed the difficulties encountered in using them due to the input limitations that mobile phones present. Granularity is also considered a key issue for this author, who emphasized the importance of using m-learning to deliver small blocks to complement "existing courses with value-added features such as alerts, personalized agents or communications aids, and access to interaction or discussion utilities that help users convert their dead-time to productive activity while in transit without access to computers and Internet." Rau et al. (Rau, Gao, & Wu, 2008) present a comparative survey of the effects of mobile mailing, discussion boards and a combination of both on the motivation, pressure experienced and learning performance of secondary students. Their findings suggest that such tools do not have a significant influence on student performance, but that SMSs decrease students' experience of pressure, in contrast to discussion forums which increment it. These two studies primarily analyzed the impact that mobile wireless communications may have on education. As for the set of *studies that analyze content and its delivery*, Huang et al. (Huang, Kuo, Lin, & cheng, 2008; Huang et al., 2010) analyzed the importance of context-awareness of mobile applications in primary education, while Rutchter (Ruchter, Klar, & Geiger, 2010) indicated its possible advantages for incidental continuous learning. The latter study compared diverse delivery methods for different age groups, and found no significant differences in environmental literacy and motivation between traditional media (textbook and personal contact with a teacher) and mobile approaches. Indeed, adults tend to prefer paper-based media or personal contact with a teacher to a PDA (Personal Digital Assistant). In the opinion of Ruchter et al. familiarity with paper is as important as the greater capacity of face-to-face communication for tailoring the presentation of content in determining such preferences. Relevant *studies on mobile assessment* include tools that personalize English vocabulary learning through questionnaires based on different pedagogical theories (Chen & Chung, 2008), and systems designed for self- and peer-assessment of classroom activities (Chen, 2010). These systems were tested with university students, and although relevant results are presented, the latter systems are only useful for specific group activities that took place during classes, while the former tool increments the cost of developing content (questions). This is a common problem which has been reported for many personalized educational systems (Brusilovsky, 1999; De Bra et al., 2003; Triantafillou, Georgiadou, & Economides, 2008), and which can be partially mitigated through standardization (Barchino et al., in press; De-Marcos, Barchino, Martínez, & Gutiérrez, 2009). In our study, we examined these issues with the aim of presenting an interoperable architecture that permits scalable development and deployment costs.

The research presented here is based on the foregoing ideas. Our aim was to build a mobile application that could be used as an aid to student self-evaluation. Teachers first designed their learning action in a traditional way, but then used the new tool as a means of support. They were thus able to provide the students with a set of questionnaires designed to reinforce learning. Using this tool, we demonstrate how current technologies enable mobile learning activities to be conducted consistent with the aforementioned trends. However, this aim alone would probably not make a real contribution to the current state of the art. Therefore, the second objective of this research was to assess the real usefulness of the application in terms of the performance observed in different groups of learners, and also in terms of students' attitudes towards the new tool and the methods that its inclusion gave rise to. None of the previously cited research included similar surveys which would have enabled us to draw any real conclusions about the true effect of mobile self-assessment on learning actions, and our aim was to fill that gap.

3. The system

A web based system was designed and built to support mobile self-assessment in traditional class-based learning. The architecture (Fig. 1) comprises three different systems: (1) a web server to store, deliver and evaluate online tests, (2) the mobile application that students

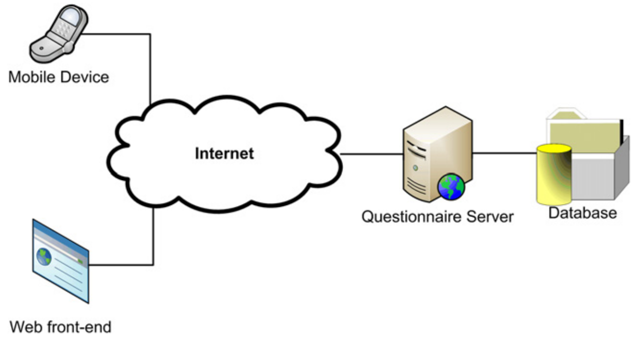

Fig. 1. System Architecture.

use to connect to the server, download questionnaires and complete them, and (3), a web based front-end that offers different functionalities to each kind of user (teachers and students). The mobile application runs on students' phones (mobile device in Fig. 1.), enabling students to access the questionnaire server through the Internet. Alternatively, students may also use the web front-end, which is a Java Server Pages (JSP) application, to access the questionnaire server using any computer with a web browser. The questionnaire server implements the functionality to: (1) detect which kind of client (web or mobile) is performing each request, (2) respond to clients' requests by serving questionnaires on demand, (3) process answers to questionnaires, show results to students and store their performance on the database, and (4) permit teachers to create, configure and review tests. Teachers must use the web application to configure questionnaires and to review students' results. An administrator's role also exists, with responsibilities related to the management of users (students, teachers and other administrators). The system also implements a database which stores information about users (students, teachers and administrators), questionnaires and questions.

The system was developed using Java technology (JME for mobile applications) and XSLT transformation sheets. This latter technology makes it easier to adapt to web and mobile system requirements. The mobile application was tested on multiple devices as part of a strategy to cover as much as possible of the available spectrum. This was essential since the end users (students) would need to be able to use their own mobile phones. Thus, through conducting as many tests as possible, it was hoped to reduce the likelihood of compatibility and run-time problems. Following the guidelines given by Mallick (Mallick, 2003) for testing mobile applications, we chose two devices from the lower end of the spectrum (namely Samsung™ J700 and Motorola™ L7), two middle of the range devices (a Nokia™ 6200 Navigator and 6110) and finally two from the upper end of the spectrum (an HTC™ TyTN II and a Toshiba™ G810). Devices from the lower end of the range are the cheapest option but they provide only the most basic features, whilst devices from the upper end of the range are usually the newest and most expensive devices offering a huge set of functionalities. Devices located in the middle of the range offer an interesting set of features at a reasonable price, and in many cases represent formerly cutting edge handhelds which have been displaced along the spectrum due to the ever-changing technological environment. A further aspect of such constant change is that it increases the difficulty of finding reliable up-to-date comparative information about devices and categories. The authors performed the selection by consulting specialized websites and vendors' home pages. Irrespective of these considerations, global requirements were that all mobile devices be java-enabled to run the application, and furthermore, they had to be compatible with current Internet connection technology (e.g. GPRS or UMTS) in order to be able to access the questionnaires located at the remote server. Software tests were run on all the aforementioned devices and the application was not released until all compatibility issues had been resolved.

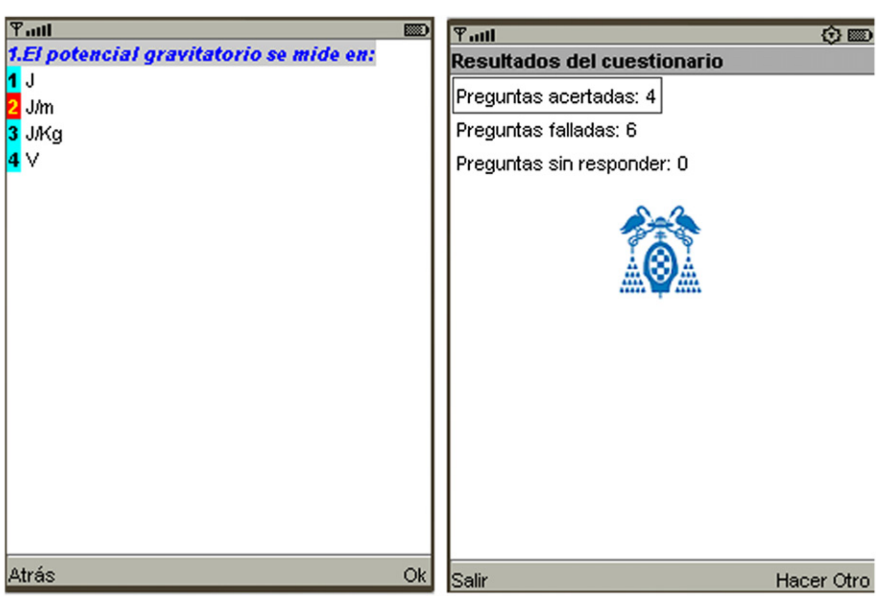

Fig. 2. Mobile Application Screenshots (in Spanish). Image on the left: A question, Image on the right: Post-test results.

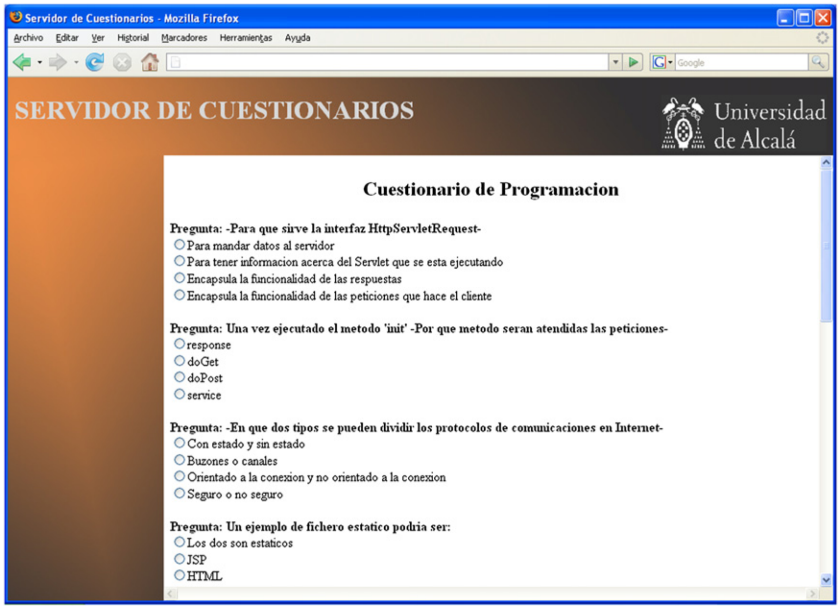

Fig. 3. Web application for the students (in Spanish).

Each student was provided with a login word and a password so that they could access both the mobile and the web application. First, they had to connect to the server, where a list of all available subjects and tests was displayed. They could complete any of the available tests, receive their results and review their answers (Fig. 2). The web application included the same features, the only difference being that all questions were presented in a row (Fig. 3).

Teachers could upload and configure tests. One important feature of the system was that it supported the IMS Question and Test Interoperability (QTI) specification (IMS, 2006): the system stored and managed all tests and questions internally using this format. QTI is a widely adopted specification that ensures interoperability between systems; thus, tests that conform to the specification can later be moved to any other compliant system. QTI specification takes into account a wide range of question types, including multiple-choice, gap filling, ordering, association and open answer amongst others. However, at present, the mobile application only supports multiple-choice questions. Therefore, teachers had to design questionnaires using only this kind of question, although the number of answers per question varied depending on how many each teacher felt were suitable. Teachers were also able to review each student's achievement, as all their personal scores were stored (Fig. 4).

Whilst the authors are aware that the features offered by the architecture and the system are limited, making it difficult to employ our system as the sole or central part of a learning action, it should be borne in mind that it was designed specifically to be used as a complementary system for incorporation into a new or currently existing learning action. Current literature supports this approach (Chen, 2010; Houser et al., 2002; Huang et al., 2010; Lehner & Nösekabel, 2002; Motiwalla, 2007; Ruchter et al., 2010; Vavoula et al., 2004). Our aim was to design a mobile system that could be integrated into current learning activities with low impact on teacher preparation times. Furthermore, the use of open software and open standards facilitates its extensibility. QTI questions and tests are stored and delivered using an XML dialect. QTI compliance ensures that tests and questions are both portable and cross-platform. They can be exported and used in other compliant systems or they can be created using other platforms and/or editors for subsequent incorporation into the questionnaire server. In addition, the use of XSLT technology facilitates easy transformation of XML data to any output (user-readable) format, such as those used by the mobile and the web applications (please note that from a technological point of view, these two latter formats are rather different). To deliver the questions stored in the system to other devices, such as PDAs for instance, would require the development of new XSLT transformation sheets in order to adapt the contents to each new media.

Huang et al. (Huang et al., 2008) indicated the common problems encountered in m-learning applications: (1) software integration, (2) limitations of the web browser, (3) interface usability, (4) reduced size of the screen, and (5) limitation of the battery life. Such limitations are of particular relevance when the application is intended to run on students' personal phones; in this case, decisions need to be taken in an attempt to alleviate the impact such issues may have. Of the problems listed above, item 2 can be mitigated by developing a mobile application that does not run on the web browser. Items 3 and 4 can be alleviated by designing an interface that minimizes the amount of information displayed and the input required from the user. This was the main reason for preferring multiple-choice questions to other kinds of questions, since these questions can usually be stated in a few lines and require the selection of one or more choices. A few mobile phone buttons can then be programmed to select/unselect each option. Moreover, various experimental studies (Chen, 2010; Ventouras, Triantis, Tsiakas, & Stergiopoulos, 2010) support the validity of this assessment method. Solving the problem represented by item 5 was beyond the scope of this study; however, students were advised to charge their devices before taking the tests and teachers were advised to design tests of no more than approximately 10 questions, in order to reduce connection times to a maximum of 20 min. Finally, item 1 was especially difficult to tackle. When the technological framework was set up, our decision was to define the minimal software requirements that handheld devices would have to meet in order to run the application. The Java Mobile Edition (JME) virtual machine was the only requirement and mobile devices had to have it. Where this is not preinstalled, a free version is available at Sun Microsystems™ web site[1]. To

[1] http://www.java.com

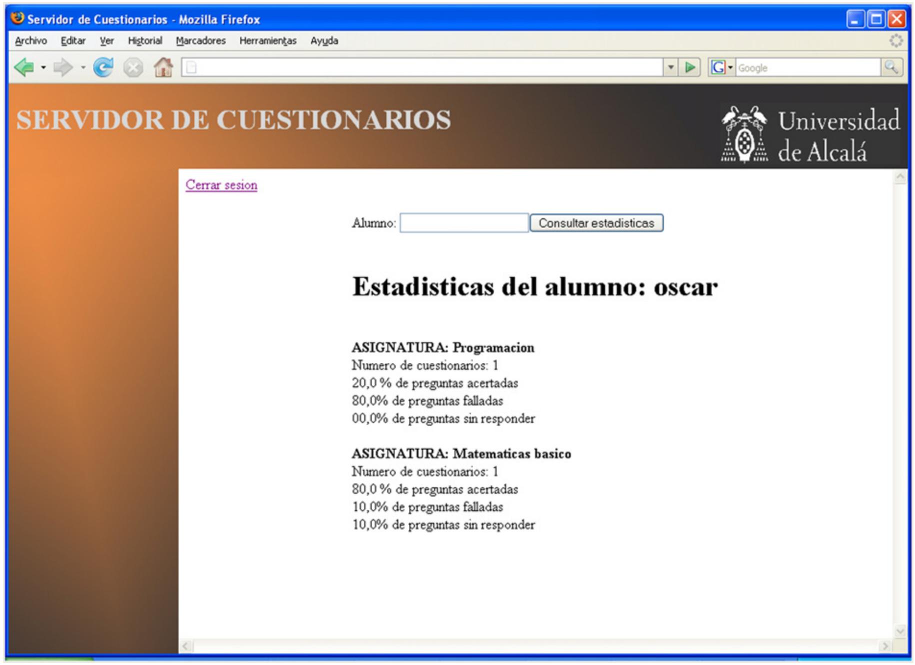

Fig. 4. Web application showing a student's achievement (in Spanish).

help those students whose mobile phones did not meet the minimum requirements, and in order to offer further assistance regarding any technical problem that might arise, support was provided as described in the next section.

4. The experiment

4.1. Learning actions and experimental groups

Our aim was to test the improvement that mobile assessment produces in student achievement at different levels. Therefore, we decided to choose three different experimental groups; two of them were attending secondary school and the third group was attending university. The first group was taken from a third year Technology course in secondary education (14–15 year old learners). The second experimental group was drawn from a Physics course from the sixth (and final) year of secondary education (scientific branch). Learner age in this group ranged from 17 to 18 years old. Finally, the third group was taken from a Nursery course from the third year of a Life Sciences degree (20–21 year old learners). These groups were selected in order to cover a wide range of ages, from adolescence to adulthood, enabling us to collect sensitive data in order to infer conclusions about the improvement that mobile assessment can produce when it is targeted at different age groups. Please note that there was a time-span of two academic years between each experimental group. Different experimental studies (Attewell & Webster, 2004; Smyth, 2004) have reported on the learning potential that new technologies may have for young people, as these latter tend to be excited by, and interested in, new technologies and such interests can be exploited to engage them in training education. Our principal aim was to assess the improvement in academic achievement and motivation that mobile assessment can produce in teenagers and young adults. Student age range was, therefore, 14 to 21 years old. Groups were selected by taking a group from the lower end of the age range (14–15 year old secondary education students) and another from the upper end of the age range (20–21 year old university students). In order to obtain sensitive data concerning the evolution of student achievement and motivation in relation to their age, the final

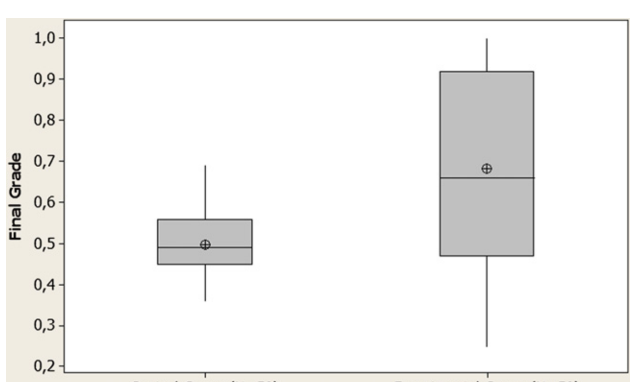

Fig. 5. Box plot of the Final Scores for the Technology Group.

Table 1
Grades for the technology group.

Measure	Control Group ($N = 50$)			Experimental Group ($N = 50$)			Significance
	Mean	Std dev	Std err	Mean	Std dev	Std err	p
TLO#1	.4667	.3227	.0833	.6867	.3647	.0942	.091
TLO#2	.6000	.2070	.0535	.7000	.3433	.0886	.342
TLO#3	.5167	.3468	.0895	.7167	.2968	.0766	.101
TLO#4	.4000	.2276	.0588	.6833	.3468	.0895	.013
TLO#5	.5067	.2251	.0581	.6267	.3369	.0581	.261
Final score	.4980	.2084	.0226	.6827	.2730	.0705	.019

group was taken from the middle of the age range, which in our case were 17–18 year old secondary students. Primary students were not included as the authors considered them too young to own mobile phones.

The size of each experimental group was 50 students from the Technology course, 48 students from the Physics course, and 28 students from the Nursery course. Our initial aim was to have 50 subjects per group. It should be noted that it was especially difficult to obtain a larger sample for university courses, even when all enrolled students were included in the study. For this reason, it was found necessary to reduce group size to 28 in the case of the Nursery course group. In the case of the Physics group, the initial size of the experimental group was 50, but two students did not complete the final assessment, and thus the group's size was finally reduced to 48. Control groups of the same size for each subject were also established: $N = 50$ for the Technology and Physics control groups and $N = 28$ for the Nursery course group. All students from each course were chosen from the same institution. One teacher was recruited for each group, to assist the researchers and to coordinate all the class activities related to this research. Teachers were also responsible for assigning students to each experimental and control group in such a manner that no significant difference in previous learning performance existed across groups. In order to fulfil this condition, achievement information from the previous semester was available for each student. Students in the experimental groups were also required to have a mobile phone, and although this may seem to be an exclusion criterion, we would like to point out that all students had one and thus all were eligible. No other background information, apart from previous achievement and mobile phone availability, was considered.

The next step was to organize the subjects being assessed in each course into a set of learning objectives (LOs). For the Technology and Physics courses, these learning objectives were taken from the official syllabus set out by the National Educational Board and which all schools must follow. The learning objectives were grouped into sets of 4 or 5 for each course as a means of facilitating the subsequent statistical analysis. The learning objectives for the Technology course were as follows:

- Objective #1 (TLO#1). To classify industries according to the phase of the industrial process in which they take part
- Objective #2 (TLO#2).To understand the main features of plastics
- Objective #3 (TLO#3).To classify plastics according to their properties
- Objective #4 (TLO#4).To understand how electrical energy is generated and distributed
- Objective #5 (TLO#5).To indentify the different kinds of electrical stations and to analyze the processes that occur within them

The learning objectives for the Physics course were as follows:

- Objective #1 (PLO#1). To understand the rules of linear and angular momentum conservation
- Objective #2 (PLO#2). To understand and apply Kepler's Laws
- Objective #3 (PLO#3). To understand and apply the Universal Gravitational Law
- Objective #4 (PLO#4). To understand the gravitational field

For the Nursery course, the learning objectives were drawn from the official syllabus approved for that degree by the National Council of Universities. The learning objectives were as follows:

- Objective #1 (NLO#1). To learn the vaccination schedule for the region
- Objective #2 (NLO#2). To understand and use the Mantoux Test

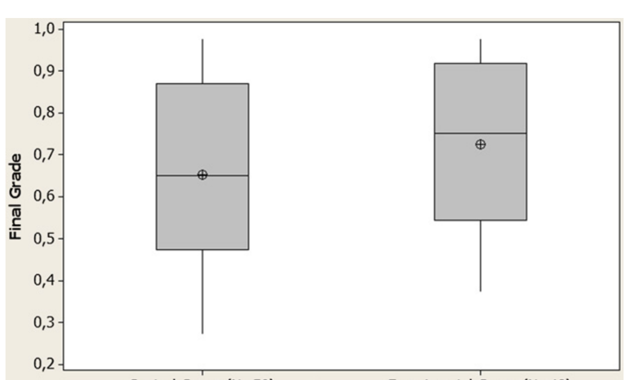

Fig. 6. Box plot of the Final Scores for the Physics Group.

Table 2
Grades for the physics group.

Measure	Control Group($N = 50$)			Experimental Group ($N = 48$)			Significance
	Mean	Std dev	Std err	Mean	Std dev	Std err	p
PLO#1	.6000	.3162	.0913	.6750	.2864	.0827	.549
PLO#2	.6583	.2109	.0609	.7250	.1815	.0524	.415
PLO#3	.7167	.2290	.0661	.7833	.2082	.0601	.463
PLO#4	.6333	.2348	.0678	.7167	.2290	.0661	.388
Final Score	.6521	.2196	.0634	.7250	.1991	.0575	.403

- Objective #3 (NLO#3). To acquire an understanding of complementary feeding for healthy 0 to 18 month old infants
- Objective #4 (NLO#4). To understand and apply diabetes treatment

Teachers designed a self-assessment test with ten questions for each learning objective. Single choice questions with 4–6 options were employed. Questionnaires were later adapted to conform to the QTI specification and uploaded to the web server. Finally, teachers and students were provided with their login word and password.

4.2. Conducting the experiments

The mobile learning tool was designed for self-assessment, so the obvious way to distribute it was to make it available to each student by installing the application on his/her mobile phone. Although this was probably the best option it must be said that there were some disadvantages to this. Firstly, technical problems may arise due to the many different kinds of devices that students have. Technical support was provided, but sometimes it proved impossible to run the application due to hardware, software or communication requirements that the terminals did not fulfil. 153 questions and problems were addressed and in 6 cases, no solution was found or provided, so 6 students (4.76% of the full sample) were not able to run the application on their phones. Secondly, the mobile application requires an Internet connection for the questionnaires to be downloaded and the response returned to the server. This obviously requires an appropriate device, but it should also be noted that this communication has a cost which, although not high, may be beyond the limited budget of a teenager or young adult. To solve these problems, teachers were temporarily provided with a set of five pre-configured mobile phones so that they could schedule different sessions in which the students were able to use these devices to perform their self-assessments. Two 50 min sessions were scheduled for each group class (14–18 students per group). During these sessions, assistance was available from teachers and also from technicians who were in attendance.

All students from the experimental groups were required to complete all the mobile tests, comprising one test per learning objective. They were free to decide whether they preferred to complete them using their own mobile phones, or using the devices provided by their teacher, or both. Finally, if they so desired, students were also able to use the web front-end to access the questionnaires from any computer with an Internet connection and a Web browser, although this was completely optional. A few of them used this method, but only after they had taken the mobile test, and usually because they wanted to recheck their answers. 32 students (64%) from the Technology experimental group completed the web tests, as did 38 students (79.17%) from the Physics experimental group and 24 students from the Nursery course (85.71%). The aim was to provide each student with a variety of ways in which to complete the tests. Students from the three control groups also took the same tests, but in this case, these were delivered in printed format on paper during the lectures.

As all the mobile sessions were intended for self-assessment, no limit was set on the number of attempts the students could make. This was considered reasonable since the mobile assessment results did not carry any weight in the student's final grade. Moreover, the web

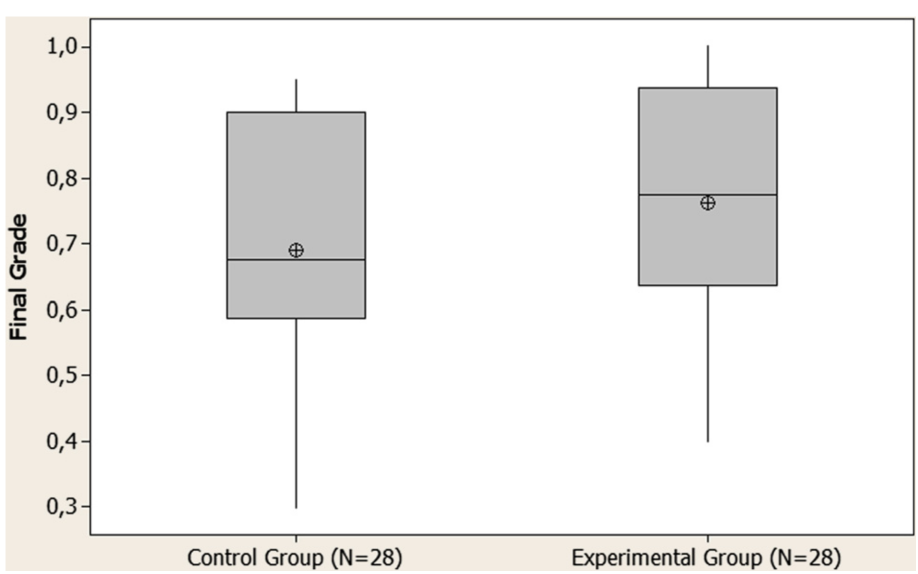

Fig. 7. Box plot of the Final Scores for the Nursery Group.

Table 3
Grades for the nursery group.

Measure	Control Group ($N = 28$)			Experimental group ($N = 28$)			Significance
	Mean	Std dev	Std err	Mean	Std dev	Std err	p
NLO#1	.6692	.2213	.0614	.7538	.2106	.0584	.328
NLO#2	.7077	.2178	.0604	.7923	.1935	.0537	.305
NLO#3	.6538	.2504	.0694	.7308	.2175	.0603	.411
NLO#4	.7308	.1974	.0548	.7692	.1797	.0499	.608
Final Score	.6904	.2050	.0569	.7615	.1802	.0500	.357

system also recorded all of a student's attempts and made this information available for teachers to use if he or she considered it worthwhile. For both these reasons, imposing any limitation on the number of attempts was considered inappropriate. Through a process of trial and error the students were able to make repeated attempts to answer correctly. Students were graded (for each module) by employing the method that each teacher normally used depending on their personal preference and experience, but also in accordance with the requirements imposed by their own institution or any other public regulations. Thus, use of the self-assessment tool did not affect the students' final grades unless they really acquired some understanding of the concepts being studied. Examination methods included papers and exams in all cases, and practical tests in the Nursery course. Final grades were also provided by the teachers; to compute them, it was assumed that each module carried the same weight. All the experiments and grading procedures were conducted during the 2008/2009 spring semester, and all the aforementioned learning objectives formed part of the course syllabus taught during that semester. It is also important to note that control groups were selected from the same institutions, taking care to choose those which had shown similar levels of achievement (up until the time of the experiment) as the experimental group. Teachers were provided with performance data from the previous semester in order to ensure that this condition was fulfilled.

5. Results and conclusions

In this section, outcome data collected from each group is presented and discussed. It should be borne in mind that each teacher provided a grade for each learning objective and each student, together with a final mark for the entire learning experience. Students' opinions were also appraised in an attitudinal survey.

5.1. Achievement improvement

The students' achievements were collated into the set of defined learning objectives for each experimental and control group. They were also normalized in the range 0–1. It should be noted that according to our national system and following this normalization, a final mark of .5 or above is a pass mark. Fig. 5 includes a box plot with a visual representation of the students' final marks for the *Technology* course. Table 1 shows the mean value, standard deviation and standard error for each learning objective and the final grade for both groups (control and experimental), together with the p-value returned by an independent-2-sample t test. Descriptive statistics show that the final mean scores of 50 students for the control and experimental groups are .498 and .6827, respectively. This represents an improvement of 18.47% over the maximum score. This study also found that the difference in the mean scores was significant for the final mark ($p = .019 < .05$). Different results were returned for each learning objective, with increments of 22% in TLO#1 ($p = .091$), 10% in TLO#2 ($p = .342$), 20% in TLO#3 ($p = .101$), 28.33% in TLO#4 ($p = .013$) and 12% in TLO#5 ($p = .261$). Although there is statistical evidence of improvement in student achievement for three modules and for the final grade (CI $\approx$ 90%), we are aware that due to the small sample size and limited functionality of our application we cannot generalize from the results.

Results for the *Physics* course are summarized in Fig. 6 and Table 2. The graph does not show any significant difference, with the exception of an increment in the median (approximately 10%). Descriptive statistics in Table 2 show that the mean scores for the control and experimental groups are .6521 and .725 respectively, which represents a moderate improvement of 7.29% over the maximum score. Similar results are returned for each learning objective. Results of the independent-samples t-test were not significant ($p > .05$ in all cases), and again, the sample size does not allow us to infer generalized statistical significance.

As for the Nursery course, the results are shown in Fig. 7 and Table 3. No significant differences could be determined between the final score of both groups in Fig. 7. Results of descriptive statistics listed in Table 3 show that the final mean scores of 28 learners for the control and experimental groups are .6904 and .7615, respectively. This represents an improvement of 7.11% over the maximum score. Similar

Table 4
Questions of the attitudinal survey. Answers were provided in a five-point likert scale (1-strongly disagree, 2-disagree, 3-undecided, 4-agree, 5-strongly agree).

#	Question
1	Questionnaires were presented effectively
2	I learned about [the course topic]
3	I enjoyed the experience
4	Using the mobile tool was easy
5	The proposed practice activities were useful
6	There was a sufficient number of exercises
7	There was sufficient time to complete the exercises
8	My level of involvement was high
9	I would like to learn more about [the course topic]
10	This was a worthwhile learning experience

Table 5
Results of the attitudinal survey in the different groups.

Question	Technology ($N = 50$)		Physics ($N = 48$)		Nursery ($N = 28$)		Overall ($N = 126$)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	3.67	.82	3.42	.79	4.00	.71	3.70	.79
2	3.33	.98	3.25	.87	3.23	1.01	3.28	.93
3	3.60	1.18	3.67	1.07	4.23	.93	3.83	1.08
4	4.20	.86	4.08	.79	4.62	.51	4.30	.76
5	3.40	1.18	3.50	.67	4.54	.66	3.80	1.01
6	3.13	1.46	3.68	.78	4.00	.58	3.58	1.08
7	3.93	1.03	4.00	.85	4.62	.51	4.18	.87
8	3.73	.88	3.68	.78	3.62	.87	3.68	.83
9	3.53	1.19	3.25	.97	3.39	.87	3.40	1.01
10	3.67	1.11	3.75	.62	4.62	.51	4.00	.90
Average	3.62	–	3.63	–	4.08	–	3.77	–

results were returned for NLO#1, NLO#2 and NLO#3, with increments of 8.46%, 8.46% and 7.7%, respectively. A more moderate effect can be observed in NLO#4; mean score increased by just 3.84%. Teachers of this course may feel, when designing future learning actions, that the mobile application contributed little or nothing to the students' achievement in NLO#4. A subsequent analysis carried out with the help of teachers on the course upheld the validity of this result: it seems that the topics covered in this learning objective are mainly practical and as such are difficult to test with a mobile assessment application. As will be discussed later, a new and interesting line of research remains open here. Table 3 also gives the comparison results of independent-samples *t*-test for the control and experimental groups. This study found that the difference between the mean scores was not significant for any learning objective or for the final mark ($p > .05$ in all cases). However, due to the small sample size and limited functionality of the application, we cannot generalize from these results.

5.2. Attitudinal survey

Students from the experimental groups were also asked to answer a questionnaire of 10 items designed to evaluate their attitude towards the learning tool and their satisfaction level. The instrument used was a questionnaire based on a five-point Likert scale with the items shown in Table 4, with all the sentences scored on a positive scale. Similar instruments have been used by other researchers (Garrido, Grediaga, & Ledesma, 2008). Results are summarized in Table 5. The average for these questions is 3.77 on the five-point scale, indicating that the students' attitude to this experience was positive. The average for the Technology and Physics groups is similar (3.62), but students on the Nursery course returned a higher value (4.02), suggesting a high rate of approval. The lowest rated statement was item 2, which is related to the students' learning. This was to be expected given that the application was designed for self-assessment and reinforcement. Another statement with a low rating was item 9, which refers to the students' motivation towards new learning. We feel that it would be worthwhile for both teachers and researchers to design new experiments and learning actions which could increase student motivation, as lack of motivation is the result of a number of aspects that are not easy to summarize. All other items were rated above 3.5. The ratings for items 4, 7 and 10 are especially significant. Item 4 demonstrates how user friendly the tool is. Students very quickly become acquainted with it and it is worth mentioning that on all courses it took longer to train the teachers than it did the students. Item 7 is related to the time available to complete the activities. Given the positive rating the tool achieved, it would appear that the learning activities and sessions were adequately scheduled with enough time to complete them. The high rating given to item 10 reflects a very positive attitude towards the learning experience.

Answer variability was low since overall SD was .93, which represents less than 1/4 of the mean. Therefore, it can be affirmed that the answers were homogeneous. To complete the analysis of the attitudinal survey, Cronbach's alpha score was computed to measure the internal consistency of the survey and the results are summarized in Table 6. All the values obtained were higher than .7; this suggests that the test items measured the same construct.

5.2.1. Conclusions and future work

A new system for m-learning which consists of a mobile application for student self-assessment, the server side and a web front-end are presented in this paper. Its conformity with current specifications is a remarkable feature of this system, as it ensures that the questionnaires designed for use with this system can later be transferred to any other compliant system. Three courses at different education levels were adapted to incorporate this tool, and different sessions were scheduled to test its usability, usefulness and performance. Student achievement as regards various learning objectives on these courses was collected for statistical analysis. Results show that there was an improvement in student achievement in all cases. On a Technology course targeted at 14–15 year old students, an increment of 18.5% was observed in more than 95% of cases. On a Physics course for 17–18 year old students, a 7.3% increment was observed in approximately 60% of the sample (29 out of 48 students). Lastly, on a Nursery course for 20–21 year old students at university, a 7.1% improvement was observed in

Table 6
Cronbach's alpha score of the attitudinal survey.

Course	Cronbach's alpha score
Technology ($N = 50$)	.88
Physics ($N = 48$)	.87
Nursery ($N = 28$)	.86
Overall ($N = 126$)	.88

18 out of 28 (65%) students. Results suggest that in some cases, the inclusion of this new tool in learning actions produces a considerable improvement in student achievement. An attitudinal survey was also carried out, and the corresponding results suggest a relatively positive attitude on the part of the students.

In addition to these results, further conclusions may be drawn. Firstly, the decreasing level of improvement that occurred in the Physics and Nursery courses must be considered, and it could be argued that this may have been related to student age. It seems that teenagers feel more at home with new technologies, and that this familiarity increases their motivation; as result, performance improves. Older students are not as motivated by the mobile application as are their younger colleagues, and this may explain their lower (but still considerable) improvement. Conversations with teachers tended to confirm this explanation, but additional research will be required to confirm it and find empirical evidence to support this point. Further research will be conducted in this area with a larger sample of learners and courses. We are aware that the number of experimental groups was not very high (just 3) and that the sample size did not provide the statistical significance required for generalization. A larger study will be required to confirm the validity of the results. Nevertheless, we consider that the flexibility of this tool should facilitate its incorporation into many other learning actions, expanding on this study and its conclusions.

If we focus on an analysis of the Nursery course, we can see that the improvement observed in one learning objective (NLO#4) was remarkably low. This was due to the fact that the principal component of this learning objective was of a practical nature, making it extremely difficult to incorporate it into a self-assessment mobile activity. Therefore, it will be necessary to seek alternative ways to design learning objectives and assessment procedures within these kinds of courses. The easiest method would be to exclude such a learning objective from the m-learning activity, but we believe that research into new mobile applications should be carried out. 3D applications in particular have shown their learning potential (Chittaro & Ranon, 2007), and 3D technology for mobile devices and its application to learning is also becoming a reality (Gutiérrez, Otón, Jiménez, & Barchino, 2008). It is also important to take into account the fact that the attitudinal survey suggests that students did not learn through using this tool. Although this result was to be expected, since the tool was designed for self-assessment and therefore used to reinforce acquired knowledge rather than to gain new knowledge, we feel it is important that this issue be given greater consideration. The low ratings obtained by items related to this aspect in the attitudinal survey suggest that students have certain expectations about learning with their mobile phones. Therefore, we believe it is important to design and analyze tools that support knowledge acquisition as well as knowledge reinforcement. Personalization is another important topic in its own right, which is also connected to knowledge acquisition tools and mechanisms. Adaptive tests and systems have been exhaustively studied (Barchino, 2005) and it is also possible to find research on m-learning adaptive tests (Triantafillou et al., 2008). The system we have presented can be extended to include these kinds of tests; this could represent a significant contribution to m-learning studies. Furthermore, if at a later date learning content inclusion were to be considered, adaptive technologies also offer a wide variety of techniques for improving learning experiences (Brusilovsky, 1996, 2004). Approaches that rely heavily on standards usually focus on learning content sequencing (De-Marcos et al., 2009) but they can be explored in greater depth in order to organize and deliver personalized questions.

Acknowledgments

This project was supported by the program of Educational Innovation & Cooperation of the Council of Communities of Castilla-La Mancha under Contract No. EM2007-004. Additional information about this project could be found at: http://www.tifyc-uah.org/tifyc/index.php?option=com_proyectos&view=proyecto&id_proyecto=64&lang=en

References

Attewell, J., & Webster, T. (2004). Engaging and supporting mobile learners. In J. Attewell, & C. Savill-Smith (Eds.), *Third annual MLEARN conference (MLEARN 2004)*. Rome: Italy, Learning and Skills Development Agency.
Barchino, R. (2005). Assessment in learning technologies standards. *US-China Education Review, 2*, 31–34.
Barchino, R., De-Marcos, L., Gutiérrez, J. M., Otón, S., Jiménez, L., Gutiérrez, J.A., et al. Assessment design: a step towards interoperability. *Computer Applications in Engineering Education*. Corrected Proof, in Press, doi:10.1002/cae.20363.
Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction, 6*, 87–129.
Brusilovsky, P. (1999). Adaptive and intelligent technologies for web-based education. *Künstliche Intelligenz, Special Issue on Intelligent Systems and Teleteaching, 4*, 19–25.
Brusilovsky, P. (2004). *Knowledge tree: A distributed architecture for adaptive learning*. International Conference on World Wide Web (www). New York, USA: ACM Press.
Chen, C. H. (2010). The implementation and evaluation of a mobile self- and peer-assessment system. *Computers & Education, 55*, 229–236.
Chen, C. M., & Chung, C. J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education, 51*, 624–645.
Chittaro, L., & Ranon, R. (2007). Web3D technologies in learning, education and training: motivations, issues, opportunities. *Computers & Education, 49*, 3–18.
De Bra, P., Aerts, A., Berden, B., Lange, B. D., Rousseau, B., Santic, T., et al. (2003). AHA! The adaptive hypermedia architecture. In *Proceedings of the fourteenth ACM conference on hypertext and hypermedia.* Nottingham, UK: ACM Press.
De-Marcos, L., Barchino, R., Martínez, J. J., & Gutiérrez, J. A. (2009). A new method for domain independent curriculum sequencing: a case study in a web engineering master program. *International Journal of Engineering Education, 25*, 632–645.
Funk, J. L. (2004). *Mobile disruption: The technologies and applications driving the mobile Internet*. London. UK.: John Wiley and Sons.
Garrido, P. P., Grediaga, A., & Ledesma, B. (2008). VisualJVM: a visual tool for teaching java technology. *IEEE Transactions on Education, 51*, 86–92.
Gutiérrez, J. M., Otón, S., Jiménez, L., & Barchino, R. (2008). M-learning enhancement using 3D worlds. *International Journal of Engineering Education, 24*, 56–61.
Houser, C., Thornton, P., & Kluge, D. (2002). Mobile learning: cell phones and PDAs for education. In *Proceedings of the internartional conference on computers in education (ICCE'02)*. London (UK).
Huang, Y. M., Kuo, Y. H., Lin, Y. T., & Cheng, S. C. (2008). Toward interactive mobile synchronous learning environment with context-awareness service. *Computers & Education, 51*, 1205–1226.
Huang, Y. M., Lin, Y. T., & Cheng, S. C. (2010). Effectiveness of a mobile plant learning system in a science curriculum in Taiwanese elementary education. *Computers & Education, 54*, 47–58.
IMS. (2006). *Question and test interoperability information model – v2.1*. IMS Global Learning Consortium.
Lehner, F., & Nösekabel, H. (2002). The role of mobile devices in e-learning – first experiences with a wireless e-learning enviroment. In IEEE. (Ed.), *Proceedings of the IEEE international workshop on wireless and mobile technologies in education (WMTE'02)*. Växjö (Sweden).
Mallick, M. (2003). *Mobile and wireless design essentials*. Indianapolis, Indiana, USA: Wiley Publishing.
Mclean, N. (2003). *The M-learning paradigm: An overview*. Royal Academy of Engineering and the Vodafone Group Foundation.
Mobilearn. (2003). *The mobilearn project vision*. The MOBILearn Project.
Motiwalla, L. F. (2007). Mobile learning: a framework and evaluation. *Computers & Education, 49*, 581–596.
Nyíri, K. (2002). Towards a philosophy of M-learning. In IEEE. (Ed.), *Proceedings of the IEEE international workshop on wireless and mobile technologies in education (WMTE'02)*. Växjö (Sweden).

Rau, P. L. P., Gao, Q., & Wu, L. M. (2008). Using mobile communication technology in high school education: motivation, pressure, and learning performance. *Computers & Education, 50*, 1–22.
Ruchter, M., Klar, B., & Geiger, W. (2010). Comparing the effects of mobile computers and traditional approaches in environmental education. *Computers & Education, 54*, 1054–1067.
Shepherd, C. (2001). *M is for maybe*. Fastrak Consulting Ltd.
Smyth, G. (2004). Educational content anytime, anywhere on any device. In J. Attewell, & C. Savill-Smith (Eds.), *Third annual MLEARN conference (MLEARN 2004)*. Rome, Italy: Learning and Skills Development Agency.
Triantafillou, E., Georgiadou, E., & Economides, A. A. (2008). The design and evaluation of a computerized adaptive test on mobile devices. *Computers & Education, 50*, 1319–1330.
Vavoula, G. N., Lefrere, P., O'malley, C., Sharples, M., & Taylor, J. (2004). Producing guidelines for learning, teaching and tutoring in a mobile environment. In *Proceedings of the 2nd IEEE international workshop on wireless and mobile technologies in education (WMTE'04)*. Bristol (UK).
Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2010). Comparison of examination methods based on multiple-choice questions and constructed-response questions using personal computers. *Computers & Education, 54*, 455–461.
Wuthrich, C., Halverson, R., Griffin, T. W., & Passos, N. L. (2003). Instructional testing through wireless handheld devices. In *Proceedings of the 33th ASEE/IEEE frontiers in education conference*. Boulder (USA).