Computers in Human Behavior xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Computers in Human Behavior

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

Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on *eye tracking* and learner subjective perception

Ana I. Molina *, Miguel A. Redondo, Carmen Lacave, Manuel Ortega

Departamento de Tecnologías y Sistemas de Información, Escuela Superior de Informática de Ciudad Real, Universidad de Castilla-La Mancha, 13071-Paseo de la Universidad, 4, Ciudad Real, Spain

A R T I C L E I N F O

Article history:
Available online xxxx

Keywords:
m-Learning
Empirical study
Eye tracking
Learning efficiency
Learner subjective perception

A B S T R A C T

Mobile device usage has become part of our daily routine. Our interest is centered on their use in teaching–learning contexts: the so-called *m-learning*. In this work we try to empirically analyze the use of these portable devices for accessing learning materials. To this end, two empirical studies have been conducted with the aim of analyzing the effectiveness of several interaction devices for supporting study tasks. In an initial experiment we compared conventional access, by means of a desktop computer, with the access through *mobile phones*. A replica of this first experiment was conducted to compare these two devices with the use of *tablet* devices. In both experiments we use several sources of information: *subjective perception* of the students, their *profiles*, their *performance* on a study task, as well as the physical evidence provided by an *eye tracker*. The results obtained allowed us to conclude that the use of devices with visualization limitations (such as mobile phones) is not suitable to access and visualize learning materials, due to the fact that they impose an additional cognitive load. The results also indicate positive perception of the use of PCs and *iPads* for studying, although the latter is considered more motivating for learners.

1. Introduction

The use of mobile devices is acquiring a greater presence. They can be used for supporting a wide variety of tasks: access to the Internet, social networks, e-mail, etc. The portability of these new devices provides benefits in multiple domains. Our interest is centered on their use in teaching–learning contexts: the so-called ***mobile learning*** or ***m-learning*** (Hashim, Ahmad, & Ahmad, 2010; Motiwalla, 2007). The main advantage of this new paradigm is the possibility to access learning materials and resources "*anytime and anywhere*" (Quinn, 2001). The benefits of using mobile devices in the classroom have been researched and proved (Churchill & Churchill, 2008; Liaw, Hatala, & Huang, 2010; Uzunboylu, Cavus, & Ercag, 2009). *Smartphones* and *tablets* are ultra portable, making them easier to carry. Students can download apps to study, tweet questions, answer polls or look up information during class, obtaining all these services instantaneously. Mobile devices are familiar to students. The use of these devices does not require technological training, does not intimidate users, and remains unobtrusive in classrooms (Nyiri, 2003). These features have the potential to attract more and more learners, at least some of whom might be more motivated by lessons if these new devices were incorporated.

However, the use of these interaction devices also presents a series of disadvantages and drawbacks, mainly related to their visualization limitations (Findlater & McGrenere, 2008; Vogel, Kennedy, Kuan, Kwok, & Lai, 2007). Thus, some of these devices are not suitable to support certain tasks as, for example, the editing of long documents (Cui & Roto, 2008) or for web searching (Jones, Buchanan, & Thimbleby, 2003), in the case of *smartphones*. Also, some studies have proved that these small devices make reading more difficult and slower (Dillon, Richardson, & McKnight, 1990; Findlater & McGrenere, 2008). In conclusion, it is clear that people use these small devices differently from how they use desktop computers, and for supporting only certain tasks.

An important question to answer in this new learning scenario is whether these small devices indeed provide an equivalent experience to more traditional full-size displays (e.g., displays on desktop computers). In this work we intend to empirically answer this research question. In this article we describe two experiments in which we evaluate the access to learning materials using different interaction devices. In this work we compare access using three types of devices: desktop computers (PCs), mobile phones and *tablet* devices. In this empirical study we consider several sources of

* Corresponding author. Tel.: +34 926295300; fax: +34 926295354.
E-mail addresses: AnaIsabel.Molina@uclm.es (A.I. Molina), Miguel.Redondo@uclm.es (M.A. Redondo), Carmen.Lacave@uclm.es (C. Lacave), Manuel.Ortega@uclm.es (M. Ortega).

0747-5632/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chb.2013.04.022

information to evaluate the learning experience: the *subjective perception of students*, *learning efficiency* (based on *time* spent on studying the materials and task *performance*) and some evidence, of physical nature, provided by an *eye tracker* device (Nielsen & Pernice, 2010). The concept of ***eye tracking*** refers to a set of technologies which monitor and record the way a person looks at a particular scene or image, and specifically in what areas they fixed their attention, for how long and in the order in which he/she visually explores the material provided. The *eye tracking* technique has been applied in various disciplines and areas of study: marketing, advertising, evaluation of user interfaces (including web pages) (Nielsen & Pernice, 2010; Poole & Linden, 2004). Recently, several authors have proposed the use of this technique to provide new empirical evidences in the study of the effectiveness of educational materials and resources (Hyöna, 2010; Mayer, 2010; Ozcelik, Arslan-Ari, & Cagiltay, 2010; She & Chen, 2009; Tai, Loehrb, & Brighamc, 2006; Tsai, Hou, Lai, Liu, & Yang, 2012; van Gog & Scheiter, 2010). We believe that there is great potential in using this new source of information (of physical nature) for assessing learning technologies. Using all the aforementioned sources of information together we can analyze the learning experiences more completely. We can complement the data provided by more subjective sources of information (for example, the learner subjective perception collected by satisfaction questionnaires) and contrast them with a more objective source of information (as that which was provided by an *eye tracking* device).

This article is structured in the following sections. In the next section we present the problem derived from the use of mobile devices to access learning materials. Section 3 describes two empirical studies performed to analyze and compare the use of several interaction devices (desktop computers, mobile phones and *tablets*) to access learning materials. In this section we also briefly review the main theories and frameworks to evaluate learning resources, including the use of *eye tracking* techniques. Finally, in Section 4 the conclusions extracted from this work and possible future lines of research are presented.

2. Problem: Access to learning materials using mobile devices

From its origins, the paradigm of *m-learning* is increasingly attracting the interest not only to educators and researchers but also to companies developing learning systems and, in general, to anyone involved in the publication of educational materials. However, regardless of the interaction device used to support teaching and learning tasks, there are certain ***design principles*** or recommendations (*guidelines*) which should be taken into account when using educational materials and resources that prove to be effective and of high quality. In this regard we highlight the contributions made by Mayer, who proposes a set of design principles from the perspective of *cognitive theory* (Moreno & Mayer, 2000). These principles are: modality, contiguity, multimedia, personalization, coherence, redundancy, pre-training, signaling and pacing (Mayer & Moreno, 2004). For example, the *multimedia principle* states that better transfer occurs from animation/pictures and narration/words than from words alone. Or the *pacing principle* states that better transfer occurs when the pace of presentation is controlled by the learner, rather than by the program. Other authors, as Sorden (2005) proposes several instructional design techniques based on *Cognitive Load Theory*. These instructional principles are identified as the goal-free effect, worked example effect, completion problem effect, split-attention effect, modality effect, redundancy effect, and the variability effect. Thus, the *modality effect* asserts that effective working memory capacity can be increased by using auditory and visual working memory together rather than using one or the other alone. The *split-attention effect* states that instruction should not be designed in such a way that would cause the learner to have to divide attention between two tasks, such as searching for information to solve a problem or reading a manual while trying to practice a software application on a computer. These design recommendations and frameworks like Mayer's Cognitive Theory of Multimedia Learning provide empirical guidelines that may help us design and use learning resources and technologies more effectively.

Teachers are not usually familiar with these recommendations, and they propose the use of new resources or interaction devices without checking whether they comply with such principles and, therefore, if they benefit or, on the contrary, interfere in the learning process. Such is the case of the use of mobile devices, which have certain drawbacks, mainly derived from the physical limitations imposed by the device itself. Thus, the limited *size* of the display dedicated to visualization force users to split the content onto different screens (Findlater & McGrenere, 2008). This involves breaching two design principles proposed by Mayer: the ***spatial** and **temporal contiguity principles*** (Ginns, 2006). These principles state that "*learning is more effective when related content (e.g., graphics and associated explanatory text) are presented simultaneously, both temporally and spatially*". *Temporal contiguity* means that corresponding words and pictures should be presented at the same time, while *spatial contiguity* means that corresponding words and pictures should be presented near rather than far from each other on a page or screen. In other words, this principle states: "*don't place a visual image on one page or frame, and then discuss it on a preceding or following page/frame without continuing to show the visual image*". However, the use of interaction devices (as *smartphones* or *tablets*) to access learning content requires, in many cases, splitting the information to display onto several screens or pages, violating this principle.

Other problems, related with the use of these new devices, arise from the *interactions* necessary to visualize learning materials. The student must *navigate* between different screens to display all related information (Morrison & Duncan, 1988), or *zoom into* more clearly visualize the content (Sanchez & Goolsbee, 2010). The use of *scroll* is also necessary in many cases, which adversely affects the understanding and assimilation of materials displayed (Sanchez & Wiley, 2009). We denote the time spent on these interactions as "*not useful*" time because it is time in which the student does not devote to studying and understanding the content displayed.

All these aforementioned issues can explain the reasons why some students are dissatisfied with certain experiences framed within the *m-learning* paradigm. It is important for us to try to understand these reasons. To this end, we have conducted two empirical studies, which we will proceed to explain in the following section.

3. Empirical studies

In this section we describe the details of two empirical studies performed by the CHICO (*Computer–Human Interaction and Collaboration*) research group of the University of Castilla-La Mancha (UCLM), in Spain. In both experiments we analyze and compare the use of several interaction devices (desktop computers, mobile phones and *tablets*), to access learning materials. According to Liaw, Huang and Chen (Liaw, Huang, & Chen, 2007) there are four elements to include and therefore must be considered in *e-learning* systems: the *characteristics of the learning environment*, *satisfaction* with their environment, their own *learning activities* and *characteristics of the student*. With this in mind, in this experiment we considered multiple entries of information that allow us to consider these four aspects.

Before starting to explain the two experiments in detail, we present a brief review of the foundations and main frameworks that address the issue of assessing the effectiveness of learning resources and systems.

3.1. Assessing the effectiveness of learning materials and resources

When introducing a new resource or interaction device in educational contexts it is necessary to adequately ***assess*** its use in order to check if it helps or, on the contrary, hampers the learning process. There are several approaches or *frameworks* used to assess learning materials and resources.

Among them we highlight those that consider the ***subjective perception*** of the student as source of information. Such is the case of the *Technology Acceptance Model* (**TAM**) proposed by Davis (Davis, 1993). TAM is one of the most widely accepted theories among information-system researchers for studying the system-acceptance behavior of users (Legris, Ingham, & Collerette, 2003). TAM is inspired by the *Theory of Reasoned Action* (TRA) (Fishbein & Ajzen, 1975). TAM was the first model to mention psychological factors affecting computer acceptance, and the model assumes that both *perceived usefulness* (PU) and *perceived ease of use* (PEU) of a new technological resource are central in influencing the individual's attitude towards using that resource. An individual's attitude is hypothesized to influence the behavioral *intention to use* a certain technology (the called *usage intentions* or UI), finally relating to actual use (Ma & Liu, 2004; Schepers & Wetzels, 2006). This framework has been applied by several authors (Liaw, 2008; Liu, Liao, & Pratt, 2009; Martins & Kellermanns, 2004; Ngai, Poon, & Chan, 2007; Selim, 2003; van Raaij & Schepers, 2008) to assess the acceptance of educational resources and systems.

Other works apply the *Cognitive Load Theory* (**CLT**) to measure the effectiveness of learning materials (Sweller, van Merriënboer, & Paas, 1998). CLT focuses on human cognitive architecture and, in particular, on the limited capacity of working memory. The rationale of CLT is that the designs of instruction impose cognitive load on learners' limited working memory and that the cognitive load in turn influences learning outcomes. CLT distinguishes between three cognitive load types that demand working memory resources during learning: *intrinsic*, *extraneous*, and *germane cognitive load* (for a detailed review see Sweller, 2005; Sweller et al., 1998). *Intrinsic load* is determined by the complexity or the so called element interactivity of the learning material (Sweller & Chandler, 1994). It is generally assumed that intrinsic load is affected only by the learning content but not by the instructional design. *Extraneous load* is defined as unnecessary information processing, which is caused by the instructional design. Extraneous load is harmful to learning, because it is not directed to schema acquisition (i.e., holding information in the mind while switching between reading a text and inspecting a picture). *Germane load* is also caused by instructional design, but contrary to extraneous load, it is beneficial for learning, because it is directed to schema acquisition by directing learner's attention towards relevant learning processes that were triggered by the design (Sweller et al., 1998). The overall recommendation is that an instructional design should reduce extraneous load (i.e., information processing hindering learning) and increase germane load (i.e., information processing supporting learning).

Finally, and in addition to the aforementioned frameworks (which are more theoretical or that use more subjective sources of information), we found a new line of evaluation of learning materials which proposes the use of ***eye tracking*** techniques for the evaluation of learning materials (Hyöna, 2010; Mayer, 2010; van Gog & Scheiter, 2010). The usefulness of this technique lies in the hypothesis that *there is a link between visual scanning behavior and cognitive activity in a given subject*. Although this relationship is not always true (we do not always think about or pay much attention to what we are viewing), it is sufficiently consistent to draw objective conclusions about the cognitive processes that cause the *fixations* and *visual scanning behavior* of the subjects. In relation to this aspect, an extensive review of the different *metrics* proposed in the literature is presented in (Poole & Linden, 2004), describing also how they are measured and their interpretation. Thus, for example, *a gaze of longer duration* (*fixations*) *on an image area* generally indicates an increased difficulty in interpreting its content. In addition, a *greater number of fixations* on a particular area may indicate an increased interest by the user in its content. We therefore see that there is great potential in using this new source of information for assessing learning materials; the measures provided by the *eye tracker* allow us to contrast and complement the other sources of information (of more subjective nature).

The following section describes the details of the empirical studies, which combine the theories TAM, CLT and the information provided by an *eye tracker* device to assess and compare experiences of access to learning materials using different interaction devices.

3.2. A first empirical study: Comparing the use of desktop computers and smartphones to access learning materials

In this section we explain the details of the first experiment conducted. The aim of this first study was to compare conventional access to study materials, by means of a desktop computer, with access through *smartphones* (in particular an *iPhone* device).

3.2.1. Research hypothesis

Some of the ***hypotheses*** posed in this study are the following:

H1: "Student *performance* will be influenced by the *device* used to access learning materials".

H2: "The *time* spent in visualizing, understanding and assimilating learning contents will be influenced by the *device* used to access them".

H3: "Student's *perceived satisfaction* and *usefulness* in learning experiences will be influenced by the *device* used to access to learning materials".

In addition to studying these hypotheses we will analyze all possible relationships that may exist between the different sources of information considered in this experiment.

3.2.2. Learning materials

Regarding ***learning materials*** provided to students, these consisted of a scheme of the agile software development methodology *Scrum* (Fig. 1). For the mobile version (Fig. 1b) it was necessary to make some modifications to the material shown, in order to adapt the content to be viewed on several screens.

3.2.3. Participants

A total of 26 subjects (20 students and five professors from the College of Computer Science (ESI) in Ciudad Real as well as a professional with extensive knowledge and expertise in *Scrum* participated in the experiment. The participation of this professional enabled us to validate and prepare the learning materials used in the experiment, as well as the knowledge assessments that students had to complete.

Before carrying out the final experiment, a *pilot test* (in which five professors participated) was performed. This test was used to refine some details of the experiment (materials supplied, questionnaires, and duration), as well as to realize eye tracker calibration testing for the various versions of learning materials

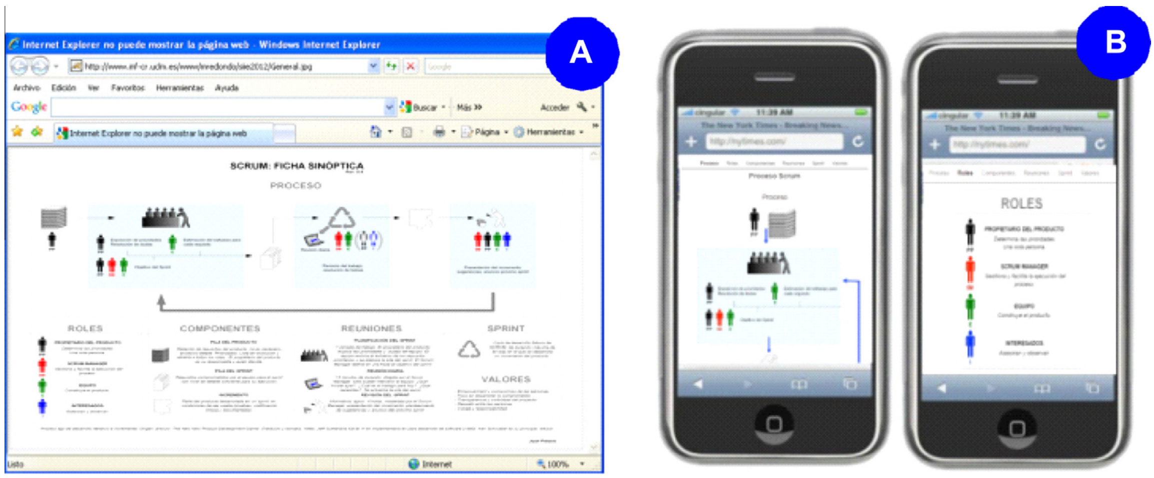

Fig. 1. Different versions of learning materials used in the experiment: (a) Desktop version (b) *iPhone* version.

(desktop and mobile version). Conducting the pilot test allowed us to prepare the *experimental task context*, that is, the conditions required to run successfully the *eye tracking* study.

As a result of this first phase, we made some decisions about the instructions that would be included in the final test, the way to supply the questionnaire and how to record the subjects' responses:

- Four of the five professors that participated in the pilot test acted as *test facilitators* during the final study. This pilot phase served as preparation of the test facilitators and as basis for the creation of a *test session script*. During the test session, the facilitator has to keep track of many things simultaneously: making sure that the participant is feeling comfortable, deciding when to intervene and taking meaningful notes during the task. Hence, most researchers recommend developing a test session script which contains everything the facilitator has to do and say to participants during the session. By using a script, the facilitator can ensure that all participants receive the same instructions and proceed through the test session in the same way. That way we can make sure that test results are not influenced by participants receiving different instructions or proceeding through the test session in different ways.
- While participants are working on their tasks, the facilitator observes what they are doing and takes notes about possible misconceptions or problems during the session. As a result of the pilot test, a form for register all these issues was created.

This first phase also allowed us to refine some aspects related to the questionnaires supplied. We made a critical review of the items of the questionnaire and modified some of them (the test size was reduced and some questions were rewritten in order to eliminate ambiguities).

The remaining participants (20 students doing a Bachelor's degree in Engineering or Computer Science), 16 men and four women, aged 19–28, voluntarily participated in this activity. The 20 students were randomly assigned to two groups: one group included those who studied the learning materials via a desktop computer (Fig. 2a and b) while the other included those who accessed the content through a mobile device (an *iPhone*) (Fig. 2c and d).

3.2.4. Equipment and eye tracker device

The experiment was performed in the *Usability Lab* of the CHICO research group of the UCLM. The laboratory includes, besides common resources in any computer lab, the proper equipment for usability and accessibility testing of interactive systems. In this sense, the lab includes *eye tracking* testing equipment, several *testing* and *interview rooms* (equipped with cameras, microphones and a PA system) and an *observation room* for monitoring tests.

The equipment used for *eye tracking* is a Tobii X60 model and the *Tobii Studio* software (version 3.0.2) for the design, implementation and subsequent analysis of *eye tracking* tests.

The *CHICO Usability Lab* has the equipment necessary to run usability tests on mobile devices. In order to do this, we use the *Tobii Mobile Device Stand (MDS)* which provides a more robust and accurate platform for conducting *eye tracking* studies on mobile devices. The MDS consists of a remote stand-alone eye tracker and a testing stand, which enables researchers to track where a participant is looking on a mobile graphical interface. The system is thought to be unobtrusive as it does not require participants to wear anything on their heads or be forced to restrain the movement of their head by the use of a chinrest.

As a result of the pilot test some *eye tracking* testing conditions, which can interfere in the session, were controlled:

- The testing room was prepared for carrying out the final test. When testing, it is important to check whether the *eye tracker* will work under the given lighting conditions before inviting any participants.
- We also prepared the testing stand (the MDS). We had to check that it enabled participants to interact comfortably with the device without obstructing the eye tracker's field of view.

3.2.5. Experimental design

Fig. 3 graphically illustrates the *experimental design* followed in this study. Each of the participants was cited at different times to participate in the experiment. The duration of the test for each subject was about 10–12 min. In the design and subsequent development of *eye tracking* tests we followed the *methodological recommendations* of Nielsen and Pernice (Nielsen & Pernice, 2009). Due to the fact that testing using mobile devices presents certain particularities we also considered the methodology proposed in (Rösler, 2012). This whitepaper provides usability professionals with a set of methodological *guidelines*, articulating how to use *eye tracking* and the Tobii MDS to evaluate mobile interfaces. This paper covers all important issues to consider in the planning, the testing, as well as the analyzing of the collected *eye tracking* data.

Please cite this article in press as: Molina, A. I., et al. Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on *eye tracking* and learner subjective perception. *Computers in Human Behavior* (2013), http://dx.doi.org/10.1016/j.chb.2013.04.022

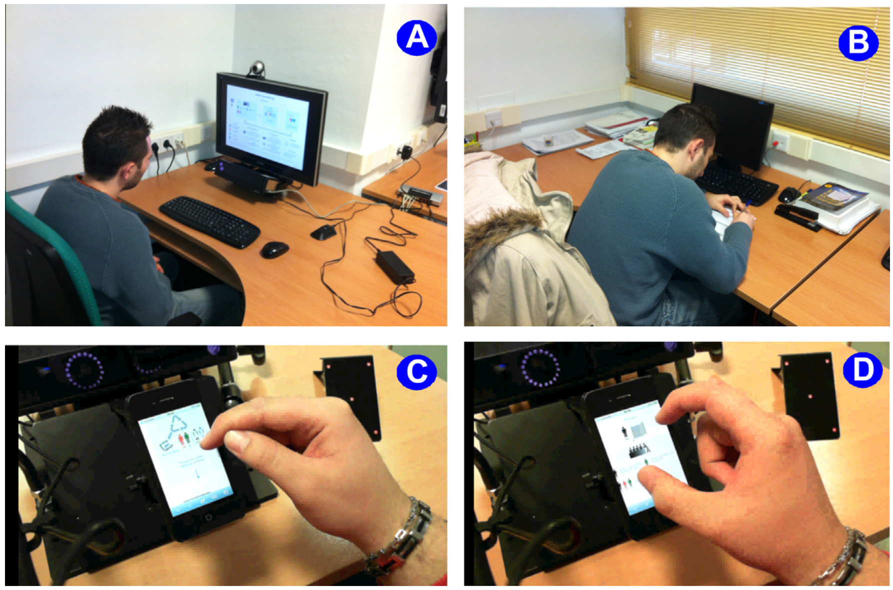

Fig. 2. Students participating in the first empirical study.

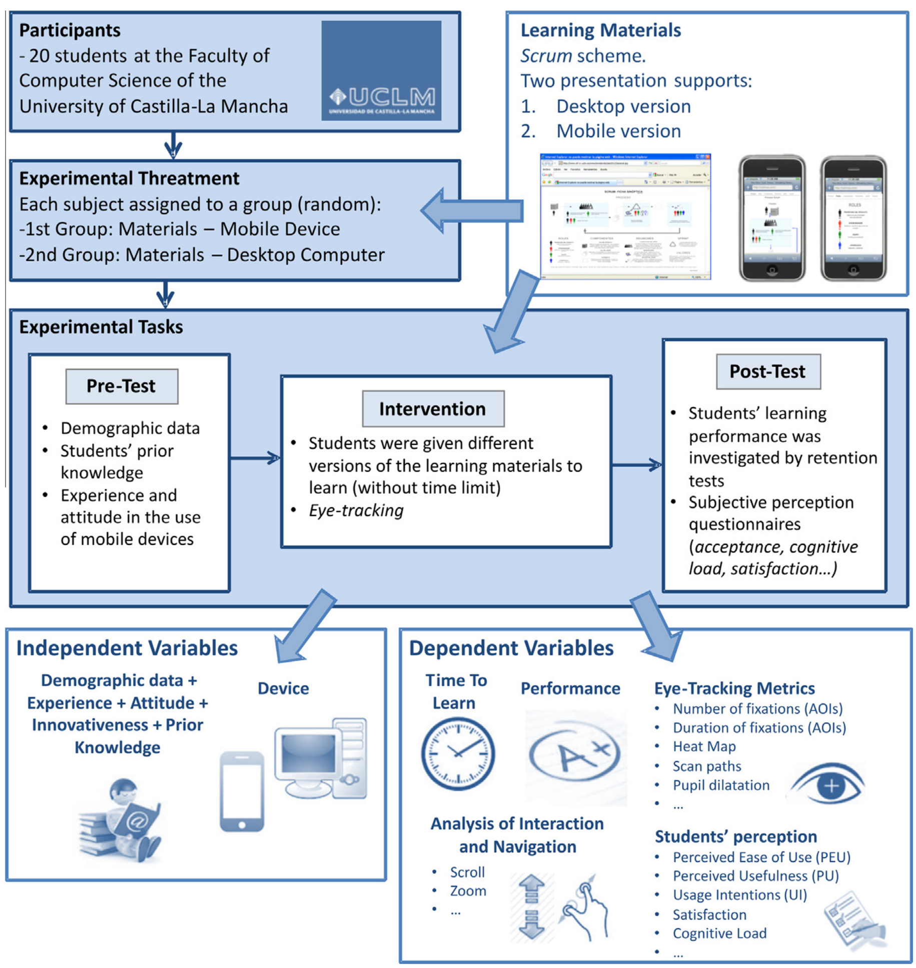

Fig. 3. Experimental design.

For each student, the test consisted of several phases (*pre-test*, *eye tracker calibration*, *intervention/study phase*, *post-test*). The objective of each of these stages, as well as the information collected in each, is described below.

Please cite this article in press as: Molina, A. I., et al. Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on *eye tracking* and learner subjective perception. *Computers in Human Behavior* (2013), http://dx.doi.org/10.1016/j.chb.2013.04.022

Firstly, the participants took a ***pre-test*** (on paper) by which researchers could gather certain data about the students' *profile*:

- ***Demographic information:*** gender, age.
- ***Personal innovativeness in the domain of information technology:*** (**PIIT** variable) (Schillewaert, Ahearne, Frambach, & Moenaert, 2005). This concept can be defined as the predisposition, attitude or tendency of a person to experiment and adopt new information technologies and/or devices. This aspect was measured by three statements that are rated on a Likert scale (1: "*strongly disagree*" to 5: "*strongly agree*"). The score of the PIIT variable is calculated as the mean of the scores of these three statements.
- ***Mobile device experience and/or expertise:*** (**MDE** variable). In this section, four sentences were included that measured the students experience using *mobile devices* (MDE1), *tactile mobile devices* (MDE2), *iPhones* (MDE3) and *m-learning* environments (MDE4). These questions were answered using a 5-point Likert scale (ranging from 1 which means "*no experience*" to 5 which means "*well experienced*").
- ***Attitude towards m-learning:*** (**AML** variable). Next, six statements were included that enabled us to determine the attitude of students regarding the use of *mobile devices* in educational contexts (AML1); the use of *mobile devices* to study (AML2); the use of *mobile phones* in educational contexts (AML3); the use of *mobile phones* to study (AML4); their preference for studying using *desktop computers* (AML5); and their preference for a *hard copy of the learning materials* (AML6). Students scored how much they agree with these six statements on a scale of 1–5.
- ***Prior knowledge on study domain:*** (**PK** variable). In order to assess *prior domain-specific knowledge* of participants, the participants were asked to rate their knowledge on a 5-point scale ranging from "*I don't know at all*" (associated with score 1) to "*I know very well*" (associated with score 5) for one statement related to *Scrum* knowledge. Following 4 questions about the *Scrum process* (**PT**: *Process Test*) and 4 relating to items or concepts in this domain (**CT**: *Concept Test*) were included. The score of these two tests was obtained by adding one point for each question answered correctly, for a maximum score of 4 on each part. The score on the knowledge pre-test about *Scrum* (**TS**: *Test Score*) was obtained from the sum of PT and CT, for a possible maximum score of eight points.

Then the participants moved onto the study task (***intervention phase***). The experiment rules and the characteristics of learning materials were explained before this phase began. No other explanation was given during the intervention. Before visualizing and studying the learning materials, the *eye tracker* was calibrated, passing by to record the study session. *Tobii Studio* software records the task and tracks the time (in seconds) taken by each student to complete the task (**TTL**: ***Time to Learn***). Each of them spent the *time of study* he/she considered necessary, with no limitation in this regard.

It is necessary to point out that some people cannot be eye tracked at all due to various reasons. *Eye tracking* technology has been developed to work well with people who have healthy eyes and normal visual acuity. Thus, for example, people wearing bifocals are typically difficult to track. Other factors can affect the *eye tracking* sessions as, for example, the ambient lighting. Also, a participant who changes his seating position during the session or the length of his eyelashes can make the *eye tracking* more difficult. In this experiment the sample population has been restricted to "*suitable eye tracking individuals*" (Nielsen & Pernice, 2009). Individuals with contact lenses, glasses and/or poor trackability are excluded from the tests. This restriction affects the sample size, because the initial number of participants recruited may decrease (Rösler, 2012).

After completing the intervention phase, all participants were asked to fill out the ***post-test***, which consisted of several sections:

- ***Retention test.*** After the study phase students had to complete a *retention test* consisting of eight multiple choice questions. These tests allowed us to measure how much information they had retained during the study phase. The eight questions were divided into two groups: four related to the *Scrum* process (**PT**) and four referred to the *concepts* and *elements* (**CT**) of that methodology. Each correct answer was worth one point, for a possible maximum score of eight points (**TS**).
- ***Technology Acceptance Method* (TAM)-based questionnaire.** The students then completed a questionnaire designed to measure their *subjective perception* regarding the *device* used to access learning materials. The questionnaire was based on the *subjective technology acceptance* questionnaires (Davis, 1993), adapting some questions for evaluating the subjective perception of students about the use of different interaction devices. The questionnaire included several statements that measured the *perceived ease of use* (**PEU**), *perceived usefulness* (**PU**) and the *usage intentions* (**UI**).
- ***Cognitive Load questionnaire* (based on CLT: *Cognitive Load Theory*).** Next, the students scored on a scale of 1–5 a series of statements designed to measure two *cognitive load* types given by the CLT (Sweller et al., 1998): *intrinsic load* (which refers to the complexity of learning certain content that depends exclusively on the actual contents to study and not of the presentation format or device used) and the *extraneous load* (depending on instructional design, that negatively interferes with the learning process of the student). The questionnaire included three statements designed to measure the *intrinsic load*, that is, the load associated with the task "*learning the Scrum process, concepts and components*" (**TD** variable: *Task demands*) and two that allowed us to measure the *extraneous load*, that is, the load derived from the demands imposed by the device used to access learning materials (**DD**: *Device demands*).
- **Intrinsic *motivation* and *performance* subjective ratings.** Finally, participants completed a set of questions designed to measure their *interest* (**INT**), *motivation* (**MOT**) and *pressure* (**PRE**) during the study activity. Also, students rated four statements designed to measure their overall *satisfaction* (**PS** variable) with the use of a particular device to learn.

The tests and ratings used to collect all the information and to measure all these variables, in a *pre-test* and *post-test,* can be consulted in Appendix A. The main results of this first experiment are discussed in the next subsection.

3.2.6. Results and discussion

In this section we analyze and interpret the information collected during this first experiment. As discussed previously it includes values obtained from different sources. Some of them are of objective nature (TTL, PT, CT, TS and metrics provided by the *eye tracker* device), while others are obtained from subjective tests (PIIT, PEU, PU, UI, TD, DD, INT, MOT, PRE, PS).

To analyze the data obtained in an *eye tracking* session it is necessary to define the so-called "***areas of interest***" (AOI) of the content to study. These areas were defined by the research team, and were used to determine if they are visible and significant to users. The *Tobii Studio* software allows us to obtain a series of quantitative metrics for each AOI defined. In the first analysis we were interested in finding the total time spent by students to visualize

Please cite this article in press as: Molina, A. I., et al. Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on *eye tracking* and learner subjective perception. *Computers in Human Behavior* (2013), http://dx.doi.org/10.1016/j.chb.2013.04.022

and study each of these two areas, and the number of times they visited each.

For the PC version, two AOIs were defined: the part of the image that represents the *Scrum* process (*AOI-Process*) and the part that shows the details of the concepts and elements of the methodology (*AOI-Concepts*). In the case of PC version the learning materials are static and take up all the space of the screen. In this case, the AOIs defined are fixed. For the mobile version, the two aforementioned AOIs were defined, in addition to another that delimitated the navigation menus (*AOI-Navigation*) at the top of the screen. In this case, the material shown has dynamic nature. When we make an *eye tracking* testing on mobile devices it is necessary to video record the session. Then, recordings need to be divided into *scenes* in order to produce visualizations and to define the AOIs. We have to create a new scene for each portion of video recording in which the graphical user interface (or image displayed on the screen) does not change. This is a laborious process. In Fig. 4a we show a capture of the video recording, highlighting the AOIs, which can change depending of the content displayed in a particular scene or instant of time.

To facilitate the interpretation of the large amount of data collected during an *eye tracking* session, several animations and representations are typically used in order to graphically summarize the *visual behavior* of a user or set of users. Some examples of static representations are *scan paths* (Fig. 4b and d) and *heat maps* (Fig. 4c and e). In the context of *heat maps*, "*hot spots*" (marked in red) show those areas to which users have paid more attention. It can be noted that this kind of representation is not very useful in the case of the mobile version; this is due to fact that the display size is very small. However, in the PC version the *hot spots* reflect those areas that are of greater interest for the students. In particular, students paid more attention to those areas showing contents previously asked about in the *pre-test* (related to temporal data, definitions, etc.). We can, therefore, conclude that they devoted more time to studying them.

The analysis of data (Table 1) allowed us to draw some interesting ***conclusions***, which we describe below. First of all, it is important to note that the ***profile*** of the participants assigned to the two groups was quite similar in several aspects: their *preference* for new technologies (**PIIT**), their *experience* with mobile devices (**MDE**) and their *attitude* towards them in learning contexts (**AML**). In this sense we have to take into account that participants were students of the Degree in Computer Engineering, with experience in the use of new technologies. Regarding their *prior knowledge* about the domain of study (**PK**), the score in both groups was quite low and very similar (1.60 and 1.70, respectively). The values of these variables (obtained in the *pre-test*) confirmed that the groups are well-balanced and that the differences that may occur

Table 1
Some results of the first empirical study.*

	PC	*iPhone*
Prior Knowledge about Scrum (PK)	1.60 (0.70)	1.70 (0.82)
Test Score (pre-test)	4.90 (1.37)	4.80 (1.40)
Time To Learn (TTL)	197.96″ (65.83)	289.56″ (53.88)
Test Score (post-test)	6.30 (0.95)	5.50 (1.02)
Learning Efficiency (LE = TS/TTL)	0.03	0.02
TAM – Perceived Ease of Use (PEU)	3.97 (0.73)	3.50 (0.96)
TAM – Perceived Usefulness (PU)	3.83 (0.85)	3.07 (0.78)
TAM –Usage intentions (UI)	3.75 (0.86)	3.05 (0.98)
CLT – Intrinsic Cognitive Load or Task Demands (TD)	2.50 (0.28)	2.47 (0.39)
CLT – Extraneous Cognitive Load or Device Demands (DD)	1.40 (0.39)	2.55 (0.98)
Interest (INT)	3.95 (0.83)	4.00 (0.88)
Motivation (MOT)	3.50 (0.85)	3.70 (0.95)
Pressure (PRE)	2.50 (1.08)	2.50 (1.18)
Perceived Satisfaction (PS)	3.90 (0.91)	3.40 (0.93)

* We show the *mean* scores and the *standard deviations* (in parentheses).

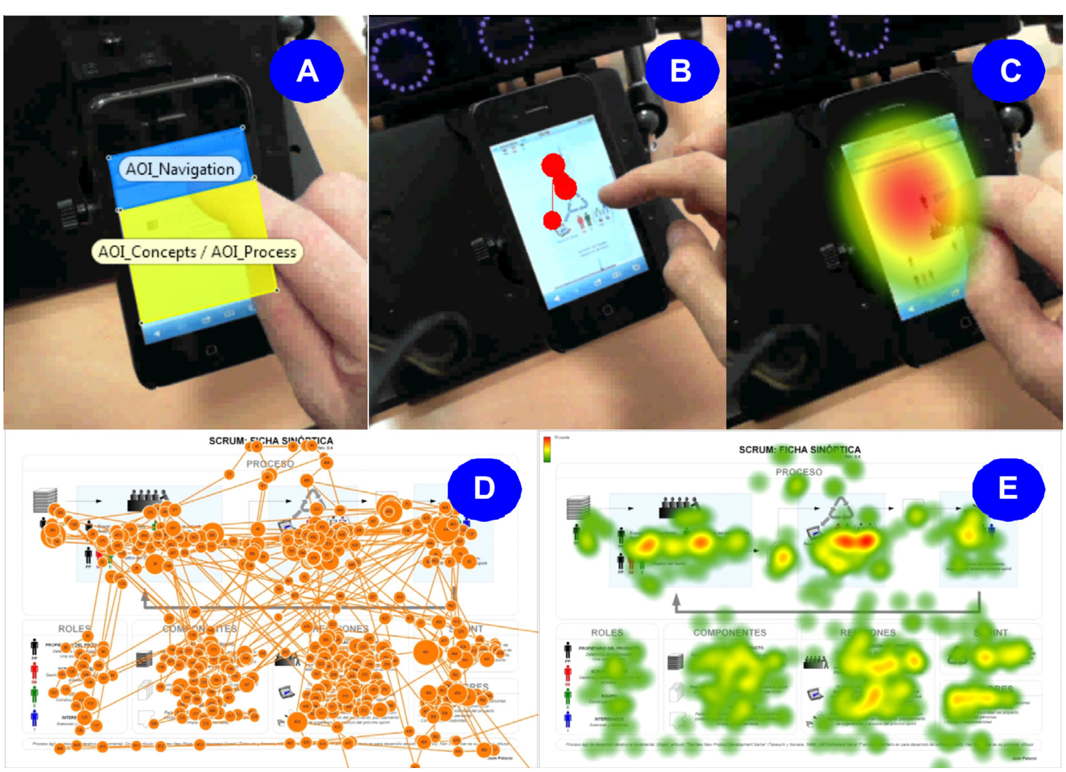

Fig. 4. Details of the *eye tracking* experiment.

as a result of the experimental task are due to the task itself and not to other factors.

Regarding the ***results of the learning task***, we begin by analyzing the average *time* dedicated to *study* by the members of both groups (**TTL**). In this sense we find significant differences. The mean TTL for members who studied on the PC was 197.96″, compared to 289.56″ dedicated by the students who used the mobile device. For the PC group the percentage of time that can be considered "useful" (dedicated to visualizing the relevant AOIs to solve the task) was 90.7% compared to 83.4% in the mobile version. In the case of the *iPhone* version the remaining 16.6% of the time was spent on interaction tasks (approximately 6.4% to navigate and the rest of the time to perform interactions such as *zoom* or *scroll*). In the case of the interaction with the *iPhone* it is necessary to indicate that several students "lost time", because they accidentally selected some functions (such as the keyboard or other menus), losing on average about 7.3″ to enable and disable these functions.

Regarding the ***performance*** of students **in the learning task** significant differences are also obtained. For students who accessed materials via the PC, the score on the questionnaires (**TS**) in the *pre-test* was 4.90 (of a maximum score of 8) passing to an average score of 6.30 in the *post-test*. For students who utilized the mobile device the improvement in the score was more moderate, from a 4.80 to a 5.50 after conducting the study session. We can see how to contiguously show the *Scrum* process diagram and the explanation of its concepts and components allowed the students to retain the content better. Considering the values of TTL and TS variables we can calculate *learning efficiency* (**LE**) (Liu, Lin, Tsai, & Paas, 2012). LE is calculated as the ratio between the score obtained by the student in the retention test and the time devoted to assimilating the contents (LE = TS/TTL). For this value, again the average ratio for the mobile version (0.02) falls below the value for PC (0.03).

Regarding the **metrics provided by the *eye tracker*** we can indicate that the average time to study the process (*AOI-Process*) by students on the PC version was 68.29″ compared to the 112.84″ spent by students who used the mobile device. As for the time dedicated to visualizing the *AOI-Concepts*, in the case of the first it was 110.30″ versus 129.40″ for students who accessed it via mobile.

As for the ***visualization patterns*** we can say that in the case of the PC version (in which the image of the process and the related explanations appeared together), most of the students alternated their attention from one AOI to another (visiting the *AOI-Process* 22 times and the *AOI-Concepts* 27 times). In the case of the mobile version, the average number of visits to the scheme of the process is much lower (being an average of three times). These scanning patterns, and study, affect retention of materials, because they impose a greater cognitive load in the case of the mobile version (Cierniak, Scheiter, & Gerjets, 2009).

These data are consistent with **subjective measures of *cognitive load*** proposed by the CLT *framework*, that we also measured in this experiment. Thus, we obtained, that in the case of the desktop version the mean scores for variables **TD** (*Task Demands*) and **DD** (*Device Demands*) are 2.50 and 1.40, respectively. The values for the mobile version are 2.47 and 2.55. As we can see the value of TD is very similar in both groups. That is, the students who worked with different versions rated the difficulty of learning the *Scrum* process and concepts similarly. There were, however, differences in terms of the load imposed by the type of device used to access the study materials (variable DD).

With regard to the ***subjective perception* measures** of the TAM *framework* (**PEU**, **PU** and **UI**), better scores are obtained for the PC version compared to the mobile version. In the case of variable PEU (*perceived ease of use*), the score for the PC was on average 3.97, versus 3.50 for the mobile. The score for PU (*perceived usefulness*) of the PC version was 3.83 compared to 3.07 for the mobile version. Finally, the UI score (*usage intentions*) was again higher for conventional access (3.75) versus mobile access (3.05).

Finally, in relation to the variable **PS**, which measured the overall student *satisfaction* with respect to the learning experience, we see that those who studied on the PC scored this aspect with a mean of 3.90, above the average 3.40 points given by the members of the other group. The values assigned to *interest* and *pressure* variables were very similar in both groups, while the score assigned to *motivation* is a bit better in the case of the use of an *iPhone* device.

To complete the analysis of the data and to infer relationships between the various dimensions considered in this study, we performed a ***correlation analysis*** (Table 2). Regarding the analysis of correlations, we detected 25 significant correlations, 16 of them positive and nine negative. We pass to comment the most relevant.

We can observe how there is a positive correlation ($r = 0.63$, $p < 0.01$) between the use of mobile *device* and the *time* necessary *to learn* the contents (TTL). Also, the *device demand* (DD), which is higher when using mobile device for studying, positively correlates with the percentage of time dedicated to navigate ($r = 0.82$, $p < 0.001$). These findings support hypothesis 2, which stated that "the *time* spent in visualizing, understanding and assimilating learning contents will be influenced by the *device* used to access them". The TTL measure also negatively correlates ($r = -0.87$,

Table 2
Correlations among study variables ($n = 20$).

	Device	TTL	LE	PEU	PU	UI	TD	DD	INT	MOT	PRE	PS	%T.Nav
Device	1.00												
TTL	**0.63****	1.00											
LE	−**0.67****	−**0.87*****	1.00										
PEU	−0.28	−0.38	0.29	1.00									
PU	−0.36	−0.41	0.35	**0.84*****	1.00								
UI	−0.37	−0.44*	0.41	**0.78*****	**0.93*****	1.00							
TD	−0.05	−0.09	0.18	−0.51*	−0.21	−0.25	1.00						
DD	**0.63****	0.36	−0.31	−**0.56****	−**0.69*****	−**0.67****	0.29	1.00					
INT	0.03	−0.45*	0.33	**0.44***	0.43	0.47*	−0.05	−0.12	1.00				
MOT	−0.12	−0.40	0.31	0.31	0.31	0.47*	0.04	−0.14	**0.63****	1.00			
PRE	0.00	0.22	−0.28	0.31	0.42	0.27	−0.26	−0.24	−0.07	−0.05	1.00		
PS	−0.27	−0.33	0.34	**0.79*****	**0.86*****	**0.88*****	−0.26	−**0.63****	0.49*	0.54*	0.32	1.00	
%T.Nav	**0.87*****	0.40	−0.40	−0.26	−0.37	−0.34	0.09	**0.82*****	0.10	0.02	−0.08	−0.28	1.00

* $p < 0.05$ (Minimum significant correlation coefficient r for sample size $n = 20$ is 0.44).
** $p < 0.01$ (Minimum significant correlation coefficient r for sample size $n = 20$ is 0.56).
*** $p < 0.001$ (Minimum significant correlation coefficient r for sample size $n = 20$ is 0.68).

$p < 0.01$) with the *learning efficiency* (LE). This performance measure is negatively related to mobile access to learning materials ($r = -0.67$, $p < 0.01$), according to the first hypothesis ("student *performance* will be influenced by the *device* used to access learning materials"). The third hypothesis, which stated that "student's *perceived satisfaction* and *usefulness* in learning experiences will be influenced by the *device* used to access to learning materials" is also tested by the correlation between the different *subjective perception* measures considered in this study. Thus, the *cognitive demand* imposed by the device (DD) positively correlates ($r = 0.63$, $p < 0.01$) with the use of mobile *device*. In relation to the *subjective perception* measures proposed by the TAM framework (PEU, PU and UI), all they are significantly and positively correlated with overall *satisfaction* (PS) of the student during the activity ($r = 0.79$, $p < 0.001$, $r = 0.86$, $p < 0.001$ and $r = 0.88$, $p < 0.001$, respectively). However all them negatively correlates with the *cognitive demand* imposed by the device (DD) ($r = -0.56$, $p < 0.01$, $r = -0.69$, $p < 0.001$ and $r = -0.67$, $p < 0.01$, respectively), which is higher in the case of the use of an *iPhone* device. This last measure (DD) also presents a negative correlation with the overall *satisfaction* ($r = -0.63$, $p < 0.01$) of the students with the use of mobile device to learn.

The most relevant correlations are graphically shown in Fig. 5 (highlighting those related to the starting hypothesis).

3.3. A second empirical study: Comparing the use of desktop computers, smartphones and tablets to access learning materials

In order to contrast the results of the first experiment with the access to learning materials using *tablet* devices, a replica was conducted. Given that the main aspects of this second study (hypothesis, equipment, experimental design, etc.) are the same of the first experiment, we will not explain all these details again. Only the differences with respect to the above study and the obtained results are discussed here.

The aim of this replica was to compare the results obtained in the previous experiment with those obtained by a new group of students who studied the learning materials using an *iPad*. The *tablet* devices share some of the advantages of mobile devices (portability, access to learning content "*anytime and anywhere*", etc.) but do not share some of its disadvantages (the size of the visualization area, necessity of navigation, scrolling and zooming for visualizing the content, etc.). These kinds of devices are more highly accepted by the students in performing some tasks (for which *smartphones* are not very suitable) like editing of documents or web searching.

In this case the ***participants*** were a set of 10 students (9 men and 1 woman) that voluntarily took part in the experiment. These subjects were undergraduate students enrolled in the fourth year of a B.S. in Computer Science at the University of Castilla-La Mancha in Ciudad Real, Spain. The average age of the participants was 22. The ***hypothesis*** posed in this replica was the same as was in the previous study (Section 3.2.1). Also, the ***learning materials*** and the knowledge domain was the same. In this case, it was not necessary to adapt the content to be viewed, because the overall *Scrum* scheme could be shown using the new device (an *iPad*) (Fig. 6). The ***experimental design*** and the test used was the same as was stated in Section 3.2.5 and Appendix A.

Next, we will discuss the obtained results. First of all, it was necessary to verify if the three groups were well-balanced. To this end, it was necessary to analyze the values obtained in the *pre-test* (related to the *profile* of the participants). The three groups were quite similar in several aspects: their *preference* for new technologies, their *experience* with mobile devices and their *attitude* towards them in learning contexts. Regarding the *prior knowledge* about *Scrum* (**PK**), the value of the new group was slightly higher than the one obtained by the other two groups (1.80).

Regarding the descriptive analysis of the rest of the data collected (shown in Table 3), we find that the average *time* dedicated to *study* by the members of the three groups (**TTL**) is different. The mean TTL for members who studied on the PC was the lower value (197.96″), compared to 289.56″ dedicated by the students who used the mobile device. The time dedicated using the *iPad* is an intermediate value between the two previous ones (252.68″). In relation to the time considered as "useful" (dedicated to study the content) the best results were obtained by the PC version (90.7%) compared to 83.4% with the *iPhone* version and 85.2% in the case of the *iPad* version. In the case of mobile devices the

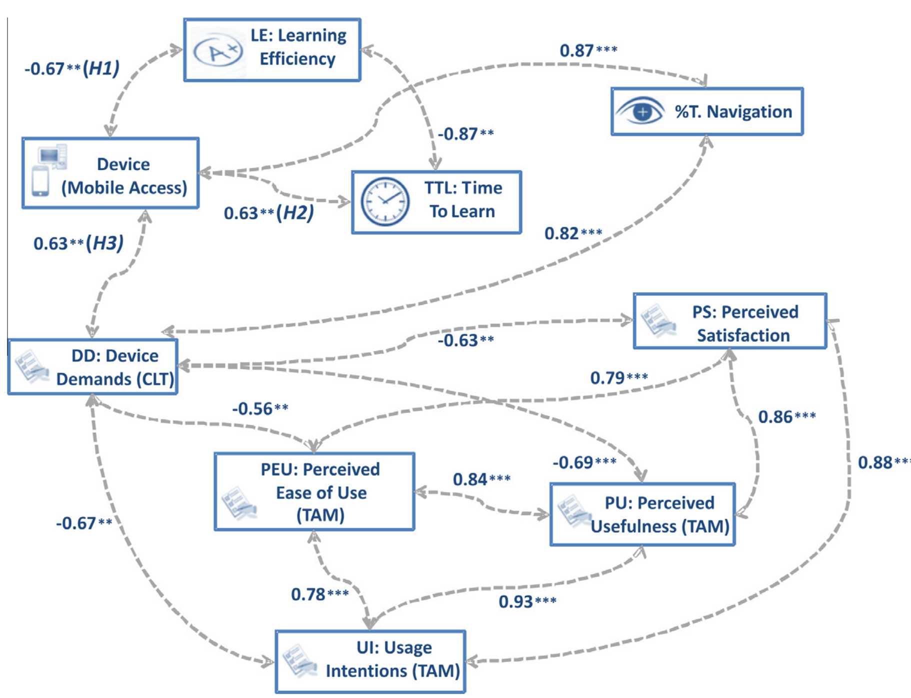

Fig. 5. Main *correlations* detected in the first experiment.

Please cite this article in press as: Molina, A. I., et al. Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on *eye tracking* and learner subjective perception. *Computers in Human Behavior* (2013), http://dx.doi.org/10.1016/j.chb.2013.04.022

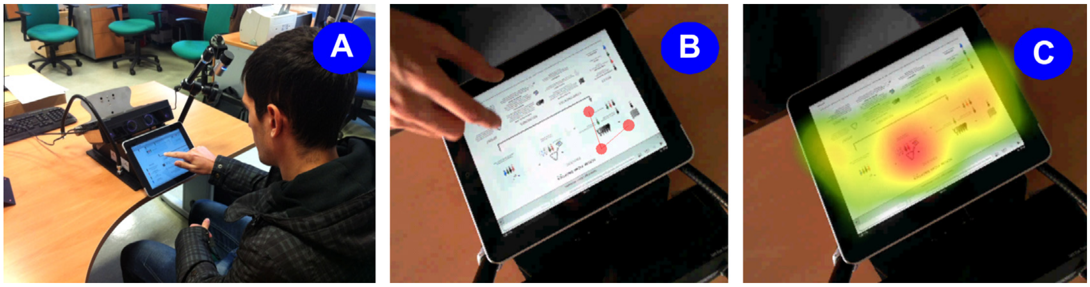

Fig. 6. Students participating in the second empirical study and examples of visual representations obtained during the *eye tracking* session.

Table 3
Some results of the second empirical study.*

	PC	*iPhone*	*iPad*
Prior Knowledge about Scrum (PK)	1.60 (0.70)	1.70 (0.82)	1.80 (0.63)
Test Score (pre-test)	4.90 (1.37)	4.80 (1.40)	4.70 (1.06)
Time To Learn (TTL)	197.96″ (65.83)	289.56″ (53.88)	252.68″ (88.15)
Test Score (post-test)	6.30 (0.95)	5.50 (1.02)	5.80 (1.32)
Learning Efficiency (LE = TS/TTL)	0.03	0.02	0.03
TAM – Perceived Ease of Use (PEU)	3.97 (0.73)	3.50 (0.96)	4.40 (0.49)
TAM – Perceived Usefulness (PU)	3.83 (0.85)	3.07 (0.78)	4.03 (0.62)
TAM –Usage intentions (UI)	3.75 (0.86)	3.05 (0.98)	3.50 (0.58)
CLT – Intrinsic Cognitive Load or Task Demands (TD)	2.50 (0.28)	2.47 (0.39)	2.40 (0.52)
CLT – Extraneous Cognitive Load or Device Demands (DD)	1.40 (0.39)	2.55 (0.98)	2.50 (0.58)
Interest (INT)	3.95 (0.83)	4.00 (0.88)	4.45 (0.37)
Motivation (MOT)	3.50 (0.85)	3.70 (0.95)	4.10 (0.99)
Pressure (PRE)	2.50 (1.08)	2.50 (1.18)	2.45 (1.17)
Perceived Satisfaction (PS)	3.90 (0.91)	3.40 (0.93)	4.03 (0.62)

* We show the *mean* scores and the *standard deviations* (in parentheses).

remaining time is spent on *zoom* and *scroll* interactions. In the case of the interaction with the *smartphone*, several students made mistakes during the study session. In the case of the *tablet* version, no student accidentally chose another function.

Regarding the ***performance*** of students **in the learning task** significant differences are also obtained. For students who accessed materials via the PC, the score on the questionnaires (**TS**) in the *pre-test* was 4.90 passing to an average score of 6.30 in the *post-test*. For students who utilized the *iPad* device the score went from a 4.70 to a 5.80. Finally, in the case of the students who used the *iPhone*, the improvement was the most moderate (passing from 4.80 to 5.50). The value of the *learning efficiency* (**LE**) was the same for the PC and *iPad* (0.03), with accessing via the *iPhone* obtaining the worst result.

The ***visualization and scanning patterns*** followed by PC and *iPad* groups were very similar. In these two cases, the students alternated their attention from one AOI to another; although in the case of *tablet* visualization the times in with each AOI is visited is lower than in the case of the *smartphone* version.

Related to the results obtained using the ***cognitive load*** ratings we can observe that the mean scores for variable **TD** (*Task Demands*) is very similar in the three groups (2.50, 2.47 and 2.40, respectively). However, there are significant differences in the value assigned by students to the *Device Demands* (**DD**). In this case, the participants considered the load imposed by the *iPhone* device (2.55) as greater than the *iPad*'s (2.50), being the PC the device that imposes the least in terms of cognitive demand.

In the case of ***subjective perception* measures** there are also significant differences. Better scores, in this second experiment, for variables *perceived ease of use* and *perceived usefulness* are obtained for the *iPad* version compared to the other two types of devices. The worst scores are assigned by the group who performed the study activity using the *iPhone.* However, the UI score (*usage intentions*) was again higher for conventional access.

Finally, in relation to the variables that measure ***motivation*** of the participants (variables INT and MOT), we can see the scores assigned to the use of the *iPad* device are greater than those assigned to other two types of devices. Also, the score of the overall ***satisfaction*** is greater in the case of the *iPad* version.

Finally, a ***correlation analysis*** was performed (Table 4). The results obtained were consistent with those obtained in the first study (Table 2). As we can see there are some variables that are related; as, for example, the measures proposed by the TAM framework (PEU, PU and UI). They all have a positive correlation with the overall *satisfaction* (PS) of the student during the activity ($r = 0.77$, $p < 0.001$, $r = 0.80$, $p < 0.001$ and $r = 0.82$, $p < 0.001$, respectively). The *satisfaction* and the *intention to use* (UI) variables negatively correlate with the *cognitive demand* imposed by the device (DD), which is higher when students use mobile devices to study ($r = -0.50$, $p < 0.01$, $r = -0.59$, $p < 0.001$, respectively). This demand positively correlates with using a mobile device to access learning materials ($r = 0.55$, $p < 0.01$). The use of this type of device negatively correlates with obtaining a greater value in the *learning efficiency* measure ($r = -0.56$, $p < 0.01$) and positively with spending more *time to learn* (TTL) the contents ($r = -0.83$, $p < 0.001$). The correlation analysis performed in this second study confirms the three hypotheses posed in this study. The most relevant correlations (highlighting those related to the starting hypothesis) are shown in Fig. 7.

3.4. A comparative analysis and discussion

In previous sections we have discussed the results obtained in the two empirical experiments. The analysis was mainly of a

Please cite this article in press as: Molina, A. I., et al. Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on *eye tracking* and learner subjective perception. *Computers in Human Behavior* (2013), http://dx.doi.org/10.1016/j.chb.2013.04.022

Table 4
Correlations among study variables (n = 30).

	Device	TTL	LE	PEU	PU	UI	TD	DD	INT	MOT	PRE	PS	%T.Nav
Device	1.00												
TTL	**0.49****	1.00											
LE	−**0.56****	−**0.83*****	1.00										
PEU	−0.24	−0.15	0.16	1.00									
PU	−0.33	−0.28	0.33	**0.76*****	100								
UI	−0.34	−0.26	0.37*	**0.69*****	**0.84*****	1,00							
TD	−0.03	−0.11	−0.04	−0.41*	−0.37*	−0.23	1.00						
DD	**0.55****	0.29	−0.35	−0.34	−**0.56****	−**0.59*****	**0.36***	1.00					
INT	0.03	−0.26	0.18	**0.50****	0.42*	0.43*	−0.02	0.01	1.00				
MOT	−0.09	−0.25	0.22	0.41	0.35	**0.46****	−0.11	−0.08	**0.54****	1.00			
PRE	0.00	0.29	−0.35	0.18	0.20	0.20	0.05	−0.02	−0.04	−0.06	1.00		
PS	−0.24	−0.17	0.24	**0.77*****	**0.80*****	**0.82*****	−0.27	−**0.50****	**0.50****	**0.62*****	0.14	1.00	
%T.Nav	**0.78*****	0.25	−0.27	−0.37*	−0.43*	−0.31	0.10	**0.52****	−0.06	−0.10	−0.04	−0.31	1.00

* $p < 0.05$ (Minimum significant correlation coefficient r for sample size n = 30 is 0.36).
** $p < 0.01$ (Minimum significant correlation coefficient r for sample size n = 30 is 0.46).
*** $p < 0.001$ (Minimum significant correlation coefficient r for sample size n = 30 is 0.57).

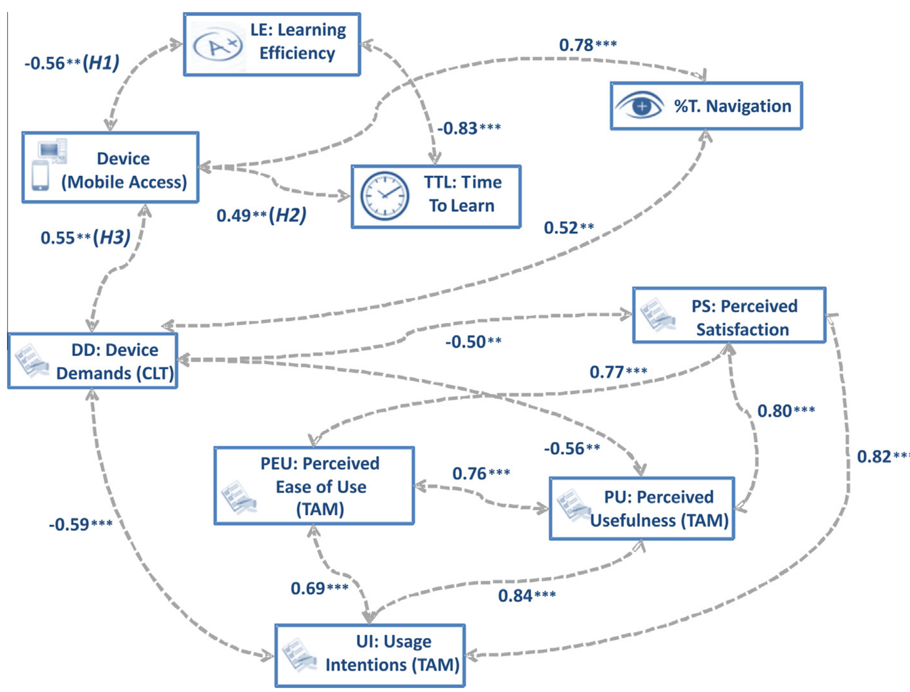

Fig. 7. Main *correlations* detected in the second study.

descriptive nature, mainly due to the limited sample size. In this section we describe a more profound statistical analysis performed to compare the results obtained in the above studies. The aim of this last phase was to analyze whether the observed differences in means in Table 3 are statistically significant and if we can extrapolate some conclusions. To do this, we consider 11 hypotheses, one for each of the factors among which we found absolute differences between their means:

$$Ho^i: \ \mu^i_{PC} = \mu^i_{iPhone} = \mu^i_{iPad}$$

Consequently, we have another 11 alternative hypotheses for each factor, which represents the existence of statistically significant differences between these values. In this context, we decided to implement the method of means contrast based on *analysis of variance* (ANOVA) for the following reasons: (a) the samples follow a normal distribution, (b) the number of groups to be analyzed is greater than two (PC, *iPhone* and *iPad*), (c) all the samples are of the same size (this being a small number: 10 subjects) and (d) the samples are independent since each experiment involved different people.

Upon analyzing the data in Table 5 we see that the factors for which we can reject the null hypothesis are TTL, LE, PEU and DD, because with a confidence level of 95% (α = 0.05), the ANOVA illustrates that the rest of the means can be considered equal, i.e. that these differences may be the result of chance. Obviously, the results are influenced by the sample size, which is small, we therefore plan to replicate the experiment with a larger number of students in the future. Thus, we accept the null hypothesis $H^4_o, H^5_o, H^6_o, H^8_o, H^9_o, H^{10}_o$ and H^{11}_o and reject H^1_o, H^2_o, H^3_o and H^7_o.

However, the analysis of variance does not indicate among which groups the means are different. To analyze this, we use the *Tukey* test, as the samples are of identical size and it allows us to compare the pairs of means. Table 6 shows the values of the function *HSD* (*Honestly Significant Difference*) for the means of factors 1, 2, 3 and 7.

Table 5
Results of the *analysis of variance* for the data in Table 3 with a confidence level of 95%. (α = 0.05).

Factor	PC		iPhone		iPad			
	μ_{PC}	σ_{PC}	μ_{iPhone}	σ_{iPhone}	μ_{iPad}	σ_{iPad}	F	p
1. Time To Learn (TTL)	197.96″	65.83	289.56″	53.88	252.68″	88.15	4.246	0.025
2. Learning Efficiency (LE = TS/TTL)	0.033	0.0082	0.02	0.0082	0.025	0.0108	6.367	0.005
3. TAM – Perceived Ease of Use (PEU)	3.97	0.73	3.50	0.96	4.40	0.49	3.593	0.041
4. TAM – Perceived Usefulness (PU)	3.83	0.85	3.07	0.78	4.03	0.62	3.117	0.061
5. TAM –Usage intentions (UI)	3.75	0.86	3.05	0.98	3.50	0.58	1.851	0.176
6. CLT –Task Demands (TD)	2.50	0.28	2.47	0.39	2.40	0.52	0.156	0.857
7. CLT –Device Demands (DD)	1.40	0.39	2.55	0.98	2.50	0.58	8.691	0.001
8. Interest (INT)	3.95	0.83	4.00	0.88	4.45	0.37	1.417	0.260
9. Motivation (MOT)	3.50	0.85	3.70	0.95	4.10	0.99	1.072	0.356
10. Pressure (PRE)	2.50	1.08	2.50	1.18	2.45	1.17	0.025	0.975
11. Perceived Satisfaction (PS)	3.90	0.91	3.40	0.93	4.03	0.62	1.577	0.225

Table 6
Results of the *Tukey* test at confidence levels of 95% and 99%.

	HSD [0.05]	HSD [0.01]	μ_{PC} versus μ_{iPhone}	μ_{PC} versus μ_{iPad}	μ_{iPhone} versus μ_{iPad}
Time To Learn (TTL)	78.55	100.71	$p < 0.05$	not significant	not significant
Learning Efficiency (LE = TS/TTL)	0.01	0.01	$p < 0.01$	not significant	not significant
Perceived Ease of Use (PEU)	0.83	1.07	not significant	not significant	$p < 0.05$
Device Demands (DD)	0.77	0.99	$p < 0.01$	$p < 0.01$	not significant

Having considered the data in Table 6, we can say that there are statistically significant differences, at 95% confidence, between the *time* spent by students who study the material using the PC and the *iPhone*, while the *time* of study with the *iPad* does not suppose a significant difference to other devices. There are also significant differences between the *learning efficiency* when using a PC versus an *iPhone.* Regarding the *perceived ease of use* (PEU), differences only exist between the *iPhone* and the *iPad*. Finally, in the case of the *cognitive load imposed* by the device (DD), with a confidence level of 99%, we found differences both between the PC and *iPad* as well as between the PC and the *iPhone*.

After this analysis, we believe that other parameters should also be considered to make a more complete study. For example, we consider interesting to include additional features of the learners, as for example, personality and individual characteristics (computer self-efficacy, openness to experience, etc.) which can influence in their acceptance and adoption of new technologies and devices (Devaraj, Easley, & Crant, 2008; Nov & Ye, 2008; Rogers, 1995). In this sense, among these individual characteristics, we also consider to include cultural differences of learners (Gabrielle & Helene, 2003; Lee, 2010; Masip, Gil, Granollers, & Collazos, 2009). Therefore, we plan to perform a replica in which we are going to take into account all these additional factors and considering a multicultural sample. However, this is beyond the scope of the work described in this article.

4. Conclusions and future works

Mobile device usage has become part of our daily routine, supporting new tasks in multiple contexts. The use of *smartphones* and *tablets* are significantly changing human–computer interaction, communication between people, and more recently learning scenarios (*m-learning*). Mobile learning has the potential to change the way students behave, the way students interact with each other and their attitude towards learning. Although the inclusion of these devices maybe will never fully replace other electronic learning approaches or classrooms, their use can complement and enrich the existing learning scenarios.

However, the use of mobile devices in educational scenarios should be carefully analyzed. It is important to try to understand the reasons why some students are dissatisfied with certain experiences framed within the *m-learning* paradigm. This line fits the present work, in which we empirically study the effectiveness of the use of new interaction devices to access educational materials. In the first experiment we compare conventional access, by means of a desktop computer, with the access through *smartphones*. A second study was performed in order to compare the results obtained in the first study with the use of a more highly accepted device to perform study activities: a *tablet* device.

In both experiments we used several sources of information: *subjective perception* of the students, their *profiles*, their *performance* on an experimental task, as well as the physical evidence provided by an *eye tracker* device. We believe that there is great potential in using *eye tracking* as additional source of information. Using this technique we can complement the data provided by subjective sources of information (the learner perception collected by questionnaires) and contrast them with an objective source of information (provided by an *eye tracker*). Using all these sources of information together we can analyze the learning experiences more completely.

The results obtained in both studies have allowed us testing the starting hypotheses considered in this research. Thus, we can conclude that learner *performance* is influenced by the *device* used to access learning materials, being more adequate the use of PC and *tablet*s. This performance can be influenced by the *time* spent in visualizing, understanding and assimilating learning contents, which is higher when we use devices with visualization limitations (mobile phones and *tablets*). The findings of the experiments suggest that the use of mobile phones is not suitable to access and visualize learning materials, because they impose additional *cognitive load*. The use of these devices requires, in many cases, splitting the information to display onto several screens or pages, violating some instructional design principles (as spatial and temporal contiguity principle). Regarding *subjective perception* of learners about the use of new interaction devices for studying (that we measure using the TAM framework), better scores for *perceived ease of use* and *perceived usefulness* are obtained for the *tablet* version compared to the other two types of devices. The worst scores are assigned to the use of mobile phones for studying. The *usage intentions* score was higher for conventional PC access. Finally, learners consider satisfactory the use of a PC and *tablets* for supporting study tasks, although the latter is considered more motivating.

In further research we plan to replicate this experiment and to obtain a larger sample of participants, which will allow us to perform a more complete quantitative analysis of the results. In these future replicas we plan to include additional features of the learners, as for example, individual and personality characteristics. Among these individual characteristics, we also consider to include cultural differences of learners.

Acknowledgements

This work has been partially supported by INTEGroup, GITE-LEARN and iColab Projects, funded by the Regional Government of Castilla-La Mancha and the Project EDUCA-Prog by the Ministry of Science and Innovation. The authors wish to thank the ESI students for their participation in this experiment.

Appendix A. *Pre-test* and *post-test* scales and items

Pre-Test
Student *profile* and *personality* traits

Gender ____
Age ____

Items (from 1 which means ''strongly disagree'' to 5 which means ''strongly agree'')	**1**	**2**	**3**	**4**	**5**
PIIT1.If I heard about a new information technology, I would look for ways to experiment with it.					
PIIT2. Among my peers, I am usually the first to try out new information technologies.					
PIIT3. I like to experiment with new information technologies.					
Items (from 1 which means ''no experience'' to 5 which means ''well experienced'')	**1**	**2**	**3**	**4**	**5**
MDE1. Experience in the use of mobile devices.					
MDE2. Experience in the use of tactile mobile devices.					
MDE3. Experience in the use of an [*iPhone*|*iPad*] device.					
MDE4. Experience in the use of *m-Learning* tools.					
Items (from 1 which means ''strongly disagree'' to 5 which means ''strongly agree'')	**1**	**2**	**3**	**4**	**5**
AML1. I think it's useful to use *mobile devices* in educational contexts.					
AML2. I think it's useful to use *mobile devices* to study.					
AML4. I think it's useful to use [*mobile phones*|*tablets*] in educational contexts.					
AML5. I think it's useful to use [*mobile phones*|*tablets*] to study.					
AML6. I prefer to use a *desktop computer* to study.					
AML7. To study I prefer to *print* the material.					
Items (from 1 which means ''very low knowledge'' to 5 which means ''very high knowledge'')	**1**	**2**	**3**	**4**	**5**
PKT1. Indicate your level of familiarity with the agile software development methodology *Scrum.*					

PIIT: Personal innovativeness in the domain of information technology
MDE: Mobile device experience and/or expertise
AML: Attitude toward m-learning
PKT: Prior Knowledge Test

Post-Test
Subjective ratings (Questionnaires based on learner subjective perception)

Technology Acceptance Method (TAM)-based questionnaire

Items (from 1 which means ''strongly disagree'' to 5 which means ''strongly agree'')	1	2	3	4	5
PEU1. Studying learning materials using this device is *easy* for me.					
PEU2. My *interaction* with this device has been flexible, direct and fluid.					
PEU3. *Overall*, I believe that this learning environment is *easy to use*.					
PU1. I think that the use of this type of device *could help me* in my learning tasks.					
PU2. Using this device enables me to accomplish study tasks *more quickly*.					
PU3. *Overall*, I find that using this device is a *useful* studying tool.					
UI1. I *intend to use* this device for studying in the *future*.					
UI2. I would *recommend* the use of this device for studying.					

PEU: Perceived Ease of Use
PU: Perceived Usefulness
UI: Usage Intentions

Cognitive load measurement (based on CLT: Cognitive Load Theory)

Items (from 1 which means ''strongly disagree'' to 5 which means ''strongly agree'')	1	2	3	4	5
TD1. The learning task was *difficult.*					
TD2. The *learning content* was difficult for me.					
TD3. The learning task required a lot of *mental activity/effort*.					
DD1. I have made an effort in order to *visualize* the learning materials (scroll, navigation, links).					
DD2. It was difficult to learn the learning materials using this device.					

TD: Task demands (Indication of *Intrinsic load*)
DD: Device demands (Indication of *Extraneous load*)

Intrinsic motivation and *performance* subjective ratings

Items (from 1 which means ''strongly disagree'' to 5 which means ''strongly agree'')	1	2	3	4	5
INT1. This activity was *fun* to do.					
INT2. I thought it was an *interesting* activity.					
MOT1. I felt *motivated* during the activity.					
PRE1. I felt *nervous* while doing this activity.					
PS1. I am satisfied with accessing learning contents using this device.					
PS2. I am satisfied with the interaction with this device for studying.					
PS3. I think that using this device for learning could be motivating.					
PS4. I like using this device for studying.					

INT: Interest
MOT: Motivation
PRE: Pressure
PS: Perceived Satisfaction

Please cite this article in press as: Molina, A. I., et al. Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on *eye tracking* and learner subjective perception. *Computers in Human Behavior* (2013), http://dx.doi.org/10.1016/j.chb.2013.04.022

Post-Test
Retention test (*Scrum*)

Mark the option that you consider correct:

PRT1. The SPRINT PLANNING a) Is the first stage of the process b) Is the final stage of the process c) is not a stage of Scrum
PRT2. After the SPRINT REVIEW a) the *Scrum* process ends b) you can return to the PLANNING of the next sprint c) the PRODUCT BACKLOG is generated
PRT3. In DAILY MEETINGS a) all the roles must be involved b) the TEAM and the SCRUM MASTER are involved c) could involve all roles
PRT4. The STAKEHOLDERS a) are involved in all three stages of the process b) are involved in the first and last stages of the process
c) are involved in the last stage of the process and possibly in the second

ECRT1. The DAILY MEETINGS have a duration of: a) 10 minutes b) 15 minutes c) 25 minutes
ECRT2. Among the VALUES of Scrum the following is not included… a) Respect among people b) Transparency and visibility of the project c) Use the least amount of resources possible
ECRT3. The SPRINT PLANNING has a duration of… a) 15 minutes b) 1 working day c) Approx. 4 hours
ECRT4. The SCRUM MASTER role a) advises and observes b) builds the product c) manages and facilitates the execution of the process

References

Churchill, D., & Churchill, N. (2008). Educational affordances of PDAs: A study of a teacher's exploration of this technology. *Computer and Education, 50*(4), 1439–1450.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split-attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25*, 315–324.

Cui, Y., & Roto, V. (2008). How people use the web on mobile devices. In *Proceedings of the 17th international conference on, World Wide Web* (pp. 905–914).

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioural impacts. *International Journal of Man–Machine Studies, 38*(3), 475–487.

Devaraj, S., Easley, R. F., & Crant, J. M. (2008). How does personality matter? Relating the Five-Factor Model to technology acceptance and use. *Information Systems Research, 19*, 93–105.

Dillon, A., Richardson, J., & McKnight, C. (1990). The effects of display size and text splitting on reading length text from screen. *Behavior and Information Technology, 9*(3), 215–227.

Findlater, L., & McGrenere, J. (2008). Impact of screen size on performance, awareness and user satisfaction with adaptive graphical user interfaces. In *Proceedings of SIGCHI conference on human factors in computing systems (CHI 2008)* (pp. 1247–1256).

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.

Gabrielle, F., & Helene, G. (2003). The effects of culture on performance achieved through the use of human computer interaction. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists (SAICSIT 2003). ACM International Conference Proceeding Series* (Vol. 47, pp. 218–230).

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*, 511–525.

Hashim, A. S., Ahmad, W. F. W., & Ahmad, R. (2010). A study of design principles and requirements for the M-learning application development. In *Proceedings of 2010 international conference on user science engineering (i-USEr 2010)*.

Hyöna, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172–176.

Jones, M., Buchanan, G., & Thimbleby, H. (2003). Improving web search on small screen devices. *Interacting with Computers, 15*, 479–495.

Lee, K. (2010). Culture, interface design, and design methods for mobile devices. In A. Marcus, et al. (Eds.), *Mobile TV: Customizing Content and Experience, Human-Computer Interaction Series* (pp 27–66).

Legris, P., Ingham, J., & Collerette, P. (2003). Why do people use information technology? A critical review of the technology acceptance model. *Information & Management, 40*, 191–204.

Liaw, S. S. (2008). Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. *Computers and Education, 51*, 864–873.

Liaw, S. S., Hatala, M., & Huang, H. M. (2010). Investigating acceptance toward mobile learning to assist individual knowledge management: Based on activity theory approach. *Computers and Education, 54*, 446–454.

Liaw, S. S., Huang, H. M., & Chen, G. D. (2007). An activity-theoretical approach to investigate learners' factors toward e-learning systems. *Computers in Human Behavior, 23*, 1906–1920.

Liu, S. H., Liao, H. L., & Pratt, J. A. (2009). Impact of media richness and flow on e-learning technology acceptance. *Computers and Education, 52*, 599–607.

Liu, T. C., Lin, Y. C., Tsai, M. J., & Paas, F. (2012). Split-attention and redundancy effects on mobile learning in physical environments. *Computers and Education, 58*, 172–180.
Ma, Q., & Liu, L. (2004). The technology acceptance model: A meta-analysis of empirical findings. *Journal of Organizational and End User Computing, 16*, 59–72.
Martins, L. L., & Kellermanns, F. W. (2004). A model of business school students' acceptance of a web-based course management system. *Academy of Management Learning and Education, 3*, 7–26.
Masip, L., Gil, R., Granollers, T., & Collazos, C. A. (2009). Multiculturalidad e internacionalización en interfaces web. *RASI, 6*(2), 191–196.
Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*, 167–171.
Mayer, R. E., & Moreno, R. (2004). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43–52.
Moreno, R., & Mayer, R. E. (2000). *A learner-centered approach to multimedia explanations: Deriving instructional design principles from cognitive theory*. Interactive Multimedia Electronic Journal of Computer Enhanced Learning.
Morrison, D. L., & Duncan, K. D. (1988). The effect of scrolling, hierarchically paged displays and ability on fault diagnosis performance. *Ergonomics, 31*(6), 889–904.
Motiwalla, L. F. (2007). Mobile learning: A framework and evaluation. *Computers and Education, 49*, 581–596.
Ngai, E. W. T., Poon, J. K. L., & Chan, Y. H. C. (2007). Empirical examination of the adoption of WebCT using TAM. *Computers and Education, 48*(2), 250–267.
Nielsen, J., & Pernice, K. (2009). Eyetracking Methodology. *How to Conduct and Evaluate Usability Studies Using Eyetracking. Nielsen Norman Group.*
Nielsen, J., & Pernice, K. (2010). *Técnicas de Eye Tracking para usabilidad Web*. New Riders: ANAYA Multimedia.
Nov, O., & Ye, C. (2008). Users' personality and perceived ease of use of digital libraries: The case for resistance to change. *Journal of the American Society for Information Science and Technology, 59*, 845–851.
Nyiri, K. (2003). *Mobile communication: Essays on cognition and community*. Vienna: Passagen Verlag.
Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*, 110–117.
Poole, A., & Linden, J. B. (2004). *Eye tracking in human-computer interaction and usability research: Current status and future prospects*.
Quinn, C. (2001). Get ready for m-learning. *Training and Development, 20*(2), 20–21.
Rogers, E. M. (1995). *Diffusion of innovations*. New York, NY: The Free Press.
Rösler, A. (2012). *Using the Tobii mobile device stand in usability testing on mobile devices*. Whitepaper.
Sanchez, C. A., & Goolsbee, J. Z. (2010). Character size and reading to remember from small displays. *Computers and Education, 55*(3), 1056–1062.
Sanchez, C. A., & Wiley, J. (2009). To scroll or not to scroll: Scrolling, working memory capacity and comprehending complex text. *Human Factors, 51*(5), 730–738.
Schepers, J. J. L., & Wetzels, M. G. M. (2006). Technology acceptance: a meta-analytical view on subjective norm. In *Proceedings of the 35th European marketing academy conference*. Athens, Greece.
Schillewaert, N., Ahearne, M. J., Frambach, R. T., & Moenaert, R. K. (2005). The adoption of information technology in the sales force. *Industrial Marketing Management, 34*, 323–336.
Selim, H. M. (2003). An empirical investigation of student acceptance of course websites. *Computers and Education, 40*, 343–360.
She, H. C., & Chen, Y. Z. (2009). The impact of multimedia effect on science learning: Evidence from eye movements. *Computers and Education, 53*(4), 1297–1307.
Sorden, S. D. (2005). A cognitive approach to instructional design for multimedia learning. *Informing Science Journal, 8*, 2005.
Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*, 185–233.
Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.
Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.
Tai, R. H., Loehrb, J. F., & Brighamc, F. J. (2006). An exploration of the use of eye-gaze tracking to study problem-solving on standardized science assessments. *International Journal of Research and Method in Education, 29*(2), 185–208.
Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers and Education, 58*, 375–385.
Uzunboylu, H., Cavus, N., & Ercag, E. (2009). Using mobile learning to increase environmental awareness. *Computers and Education, 52*, 381–389.
van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*, 95–99.
van Raaij, E. M., & Schepers, J. J. L. (2008). The acceptance and use of a virtual learning environment in China. *Computers and Education, 50*, 838–852.
Vogel, D., Kennedy, D. M., Kuan, K., Kwok, R., & Lai, J. (2007). Do mobile device applications affect learning? In *Proceedings of the 40th Hawaii international conference on systems sciences*.

Please cite this article in press as: Molina, A. I., et al. Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on *eye tracking* and learner subjective perception. *Computers in Human Behavior* (2013), http://dx.doi.org/10.1016/j.chb.2013.04.022