Computer based testing using “digital ink”: Participatory design of a Tablet PC based assessment application for secondary education

Panagiotis Siozos *, George Palaigeorgiou, George Triantafyllakos, Theofanis Despotakis
Multimedia Lab, Computer Science Department, Aristotle University of Thessaloniki, P.O. Box 888, Thessaloniki 54124, Greece

**A R T I C L E   I N F O**

Article history:
Received 30 July 2008
Received in revised form 30 November 2008
Accepted 14 December 2008

Keywords:
Human–computer interface
Interactive learning environments
Secondary education

**A B S T R A C T**

In this paper, we identify key challenges faced by computer-based assessment (CBA) in secondary education and we put forward a framework of design considerations: design with the students and teachers, select the most appropriate media platform and plan an evolution rather than a revolution of prior practices. We present the CBA application ‘MyTest’ which was developed using the participatory methodology We!Design, with the collaboration of 31 students and teachers. The application is targeted for the Tablet PC platform, provides “digital ink” capabilities and supports both closed-type and open questions, facilitating the transfer of traditional assessment practices to CBA. Both students and teachers were excited about the design sessions, and they asserted that they would rely more on educational software designed using this approach. The comparison of the ‘MyTest’ application with an assessment application developed with the same participatory methodology and the involvement of 40 undergraduate students revealed dissimilar expectations and needs of high school and undergraduate students that are often disguised or misinterpreted. A pilot evaluation of the application in real learning conditions was conducted with 31 students using Tablet PCs and with 37 students using traditional PCs. Interestingly, the Tablet PC platform rendered the ‘MyTest’ application more useful and usable to the students, validating our claims.

**1. Introduction**

Assessment is generally recognized as a key element in learning, as it empowers, directs and motivates students and provides success criteria against which to measure their progress. Given its educational significance, the emergent socio-political needs for the certification of knowledge and the growing assessment logistics due to the increasing number of students, the computerization of the assessment process appears to be an essential educational advancement; it automates and facilitates the lengthy and tedious procedures involved in the design, delivery, scoring and analysis of assessment (Sim, Holifield, & Brown, 2004). In many studies, teachers seemed to embrace this view and were positive towards computer-based assessment (CBA) specifying as its main advantages the reduction of testing and marking time, the easier scheduling and administration of assessments, the speed of results, its increased objectivity and security, the monitoring of students’ performance and its suitability for distance learning (Thelwall, 2000). Students also appeared to consider CBA as being more promising, credible, objective, fair, interesting, fun, fast and less difficult or stressful, while they stated that they preferred computerized versus written assessment (Croft, Dawson, Dawson, & Ward, 2001; Sambell, Sambell, & Sexton, 1999).

Despite the above-mentioned arguments and the expanding use of CBA in higher education, its application in the secondary education environment remains limited (He & Tymms, 2005). Firstly, the pedagogical soundness of CBA is challenged since it is usually understood to be a set of multiple choice and short answer questions, while at the secondary education level, multiple choice tests are not in frequent use. Secondly, the heterogeneous technological literacy of students of younger age poses a major threat to CBA validity due to the variance introduced into the assessment by computer related psychological states and traits. Teachers also have established beliefs regarding teaching and learning, through which they filter and interpret any change to the classroom setting (Niederhauser & Stoddart, 2001) and are often unwilling to adopt computerized testing and to change their core practices. Finally, the international software vendors produce CBA applications that are targeted primarily for higher and further education institutions (Conole & Warburton, 2005) and are often incongruent with the various technological, social and cultural particularities of different educational environments.

* Corresponding author. Tel./fax: +30 2310 998443.
E-mail address: psiozos@csd.auth.gr (P. Siozos).
As computers continue to penetrate all aspects of education and everyday life, paper-and-pencil assessment becomes disassociated from current learning practices in secondary education. In this way, young learners are not prepared to face future computer-based tests, which on many occasions will be of decisive importance in their academic placement and development. According to Bennett (2002) “as students come to do the majority of their learning with technology, asking them to express that learning in a medium different from the one in which they routinely work will become increasingly untenable, to the point that much of the paper testing we do today will be an anachronism.”

In the following paragraphs we identify key challenges faced by computerized testing in secondary education, put forward a framework of design considerations for the introduction of CBA applications into secondary education, present a CBA application based on this framework, and discuss its evaluation from multiple perspectives. The work presented is part of an extensive Greek project that deals with the creation of a technology-driven teaching classroom (“Future Classroom”) for supporting learning in secondary education.

2. Computerized assessment challenges for secondary education

2.1. Challenges introduced by the digital nature of assessment: the test-mode effect

One of the main issues discussed when it comes to the adoption of computerized testing is the equivalence between computer-based (CB) and paper-and-pencil (PP) tests (Clariana & Wallace, 2002; McDonald, 2002; Pommerich, 2004). Researchers note that these two modes of assessment are not equivalent due to the phenomenon known as “test mode effect”, in which various individual characteristics create a qualitatively different assessment experience for each subject (Clariana & Wallace, 2002). These characteristics introduce construct-irrelevant variance into the assessment by generating different cognitive and physical responses in each student which, ultimately, affect the experiential validity and accuracy of the assessment (Huff & Sireci, 2001). Examples of such unintended and irrelevant constructs include sensory capabilities, like sight and hearing, physical capabilities, like holding a pencil and typing, cognitive capabilities, like memory and knowledge, and psychological states, like attitudes and anxiety.

Although every form of assessment is more or less susceptible to construct-irrelevance, CBA in secondary education has to address, in addition, the students’ intense inhomogeneity concerning their attitudes, capabilities, prior interactions, expectations and needs towards technology (Palaiologou, Siozos, & Konstantakis, 2006). Today’s adolescents are the early adopters of modern technological developments; however the usual sweeping reference to the “Net Generation” usually conceals essential differences amongst young learners (Palaiologou et al., 2006; Volman, van Eck, Heemskerk, & Kuiper, 2005). This inhomogeneity produces latent constructs that infiltrate computerized tests and create non-equivalent assessments among the students. For example, slow typing can be a major cause of anxiety for students, having debilitating effects on assessment performance: students with less computer experience write shorter essays and achieve lower scores (Horkay, Bennett, Allen, Kaplan, & Yan, 2006).

2.2. Challenges introduced by the hardware being used: the platform effect

Closely related to the test-mode effect is the impact of the platform being used for the administration of assessments, which we may call the “platform effect”. Differences in the technical characteristics of the various computer platforms, such as monitor size, screen resolution, keyboard layout, and connection speed, introduce another source of construct-irrelevant variance (Huff & Sireci, 2001) and affect the assessment experience and the performance of students (Bennett, 2002).

The platform effect mostly influences reading and writing, which are both pivotal skills for assessment. Muter (1996) referred to no fewer than 29 variables that affect text legibility on screen and paper. Generally, children have found questions presented on computers to be more difficult than when presented on paper (Johnson & Green, 2004), an issue that would likely be exacerbated when the questions involve long reading passages. Powers and Potenza (1996) concluded that understanding test questions from laptop screens was more difficult due to the smaller screen resolution that required more scrolling. Lessen (2006) reviewed several studies which indicated that reading performance on computer monitors was comparable to paper only when higher screen resolutions (at least 1024 × 768) and monitor sizes (“17”) were used; otherwise users experienced diminished reading performance and more visual fatigue. Even different keyboards can have significant effects on writing performance e.g. the quality of undergraduate students’ essays written on laptops was lower due to the smaller keyboard (Powers & Potenza, 1996).

2.3. Challenges introduced by the type of questions supported

Most widespread CBA systems are based on objective tests and use various closed-type items, like true-false, multiple choice, fill-in-the-blank questions etc. Recent work promotes the introduction of innovative question types such as assertion-reason multiple choice items, drag-and-drop items, problem-solving simulations etc., which are more interactive and promise a more “authentic” assessment of examinees’ knowledge, skills, and abilities (Huff & Sireci, 2001). Although researchers claim that objective tests are capable of assessing cognitive skills up to the fourth level of Bloom’s taxonomy (Conole & Warburton, 2005), two significant issues hinder this prospect: adequate objective tests presuppose special training, experience, creativity and a lot of effort, shifting the burden of computerized testing back to the teacher while more complex computer operations and interactions may intensify the test–mode effect.

In secondary education, in particular, inadequate objective tests may lead to rote memorization of facts and discourage students from understanding their relationships and from exploring multiple interpretations. Divergent thinking and production of responses is precluded and students start employing “random guessing”, “reverse engineering” and similar strategies to look for the best possible match among limited given choices (Martinez, 1999). Objective tests are also regarded as inadequate when it comes to assessing discursive subjects, such as literacy and essay composition, where the creation of written or spoken language is indispensable (McKenna, 2001). Several studies have correlated the quality of student learning with the cognitive demand level of the assessment tasks (Gibbs, 1999), demonstrating that CBA may even undermine learning.
2.4. Challenges introduced by the feedback and scoring processes

Feedback is an important element in the learning process in relation not only to a learner’s academic performance, but also internal processes such as their knowledge and beliefs, goal-setting processes, strategy use and self-regulation (Butler & Winne, 1995). There is an interesting shift in students’ perceptions from performance-oriented assessment to qualitative feedback-oriented assessment. Secondary education students believe that the main purpose of assessment is to provide feedback that can be used to help them improve (Peterson & Irving, 2008), while they specify as two of the most important goals of grading, summarising achievement and increasing interest and motivation (Zeidner, 1992).

Initially, CBA was promoted as a very promising approach mostly due to the provision of automated scoring and personalized feedback without burdening the teachers (Alexander, O’reilly, Sweeney, & Mcallister, 2002). However, this automation is practical only for objective tests and presupposes substantial initial investment by the teachers in order to be meaningful. Essay and open questions, which are considered highly-valued components of effective assessment programs, are rarely used in CBA due to the need for students’ and teachers’ typing. Although research in automated essay scoring (AES) techniques is advancing, they still show low scoring credibility and do not seem likely to expand into high stake tests for quite some time (Thomas, 2003). Short comments on grades or simply confirming correct answers has little effect on subsequent performance. Irrespective of the feedback system’s sophistication, CBA does not yet seem able to completely replace teachers’ comprehensive ability to create highly personalized assessment feedback.

2.5. Challenges introduced by the homogenization of software

In traditional educational software design, experts (designers, researchers, and instructional experts) are supposed to know what students need and only seek technical solutions to address those needs. Although the active participation of teachers in the design process is expected to have a significant positive impact (Niederhauser & Stoddart, 2001), they have not been involved in the design of many educational applications. Students have even less opportunities to influence the educational software which they are invited to use. Despite being the end users of CBA applications, they remain out of the design process. Adults have their own prejudices and assumptions and bring into the design their opinions about what and how students should learn and the software being used should look like (Druin, 2002). Furthermore, international assessment software vendors are obliged, due to financial constraints, to unify the requirements of the different educational environments and do not produce tailor-made applications.

Inevitably, these views have a major impact upon the quality and the acceptance of educational software and yield products that are dictated, or at least severely constrained, by the inherent intentions of the designers (Squires & Preece, 1999). Priorities are determined without analyzing needs and difficulties faced by the teachers in their everyday classroom practices and in the various educational environments of different countries (Hinostroza & Mellar, 2001). However, the view that one assessment fits all – students, instructors, purposes, environments – seems unrealistic. Assessment software must be adapted to students’ and teachers’ characteristics, as well as to the particular purposes and environments of the assessment (Birenbaum, 1996).

3. Responding to the challenges

The introduction of CBA into secondary education requires a significant shift in students’ and learners’ practices. We embraced the view that educational transformation must involve the people and not be imposed upon them. Change is not an instantaneous incident but a complex and subjective learning/unlearning process for all concerned (Scott, 2003). There are already several CBA initiatives, such as computer adaptive testing, which due to cost, technical complexity, the need to maintain large item-banks and many other reasons have moved no further (Bennett, 2002). Aiming at the smooth integration of CBA in secondary education, we followed three design principles that meet the aforementioned challenges and take change management guidelines into consideration:

- Students and teachers should take responsibility for co-designing the assessment applications, in order to fulfill their expectations, address their deficiencies, adapt them to their computer experience levels and create products that are specific-tailored to their socio-cultural and educational traditions.
- The equipment selected for the assessment delivery should minimize the influence of the platform-mode and test-mode effects, should provide familiar and easy-to-learn interactions and should provide opportunities of use in real learning conditions.
- Application design should not cancel but enhance previous hand-written assessment practices, enriching them with new possibilities. They should be an evolution that emphasizes educational effectiveness and not a revolution highlighting technological innovations.

3.1. CBA design using participatory methodologies which involve students and teachers

Participatory design methods involve students and teachers directly in software development, based on the assumption that “as a result of their extensive experience with common educational tasks, such as information gathering, [students] (1) are able to easily recall, state and elaborate on their prior problems and needs, (2) have unconsciously or deliberately thought of and formed solutions and proposals concerning those educational processes, (3) are willing to collaborate with their colleagues on engineering joint solutions to their problems and, consequently, (4) may produce numerous diverse ideas for the construction of prototypes in a short amount of time” (Triantafyllakos, Palagiaiorgiou, & Tsoukalas, 2008). Druin (2002) argues that children can be equal stakeholders in the design of new technologies, contributing experiences and views that – by definition – are closer to their own way of thinking and acting. Given the dominance of constructivist approaches to learning, keeping students out of the design of educational software creates a paradox: while we seek students’ active involvement in learning, we decrease the necessary freedom for them to make decisions about their own learning (Squires & Preece, 1999).

By participating in the design process, teachers are also empowered and motivated to convey their needs and consolidate their pedagogical and organizational knowledge into practical assessment solutions. Early participation, involvement and cooperation are important.
factors in making them feel like co-owners of the assessment application and responsible for its successful incorporation in the learning environment. Participatory design bridges the gap between teachers’ and students’ differing perceptions of the effective characteristics of learning environments and leads to improved design and effectiveness (Könings, van Zundert, Brand-Gruwel, & van Merriënboer, 2007).

3.2. Selecting an adequate platform for assessment delivery

When choosing the assessment delivery platform, greater emphasis should be given to reading and writing performance. A promising, practical and affordable alternative to PCs has appeared with the emergence of the Tablet PC. Tablet PCs are essentially laptop computers with “digital ink” capabilities. They have screens equipped with digitizers that allow users to write directly onto them using a stylus, either to input handwritten text, or to simply operate the computer software, replacing the usual keyboard and mouse. Students are able to create clear and readable handwritten text, sometimes even easier than using paper and pencil. Built-in dictionaries recognize letters and proactively suggest complete words, rendering the process of handwriting even faster. Although Tablet PCs have smaller monitors, compared to ordinary PCs or laptops, reading on them is more efficient and almost comparable to reading from paper, since users can hold it closer to them, much like a book (Leeson, 2006). A Tablet PC is less obtrusive – compared to a laptop – as it does not have to be open in front of the user to impede direct visual contact and does not produce disruptive keyboard sounds (Willis & Mertschin, 2004). Even the natural position of the student sitting and writing at his desk need not be changed. Tablet PCs also revive several manual reading aids such as using a finger or a pen to pinpoint the reading position, enhancing the reading performance (Leeson, 2006).

With Tablet PCs, students need not be transferred to a computer lab to take a test. Instead, Tablet PCs can be moved anywhere required (e.g. within the school or even outdoors, at a museum or a gallery), thus freeing students from physical space barriers and allowing them to participate in assessments wherever learning is taking place. A faculty might also have a range of motivations for incorporating a Tablet PC in their classrooms such as capturing lesson content, providing handouts for students, increasing active learning etc.

3.3. Do not cancel but enhance previous practices, emphasizing educational effectiveness

As mentioned, teachers have established pedagogical and epistemological beliefs, through which they evaluate and adopt changes to classroom settings. From the teacher’s point of view, the most meaningful way to test the utility of CBA may be to compare it with what existed before (Scott, 2003). Teachers will not engage in transformative initiatives unless they see it as being effective, desirable and feasible.

Increasingly, pluralistic assessment programs, which incorporate multiple assessment formats, are recommended (Birenbaum, 1996). Closed-type questions allow for the quick and easy set up of assessments and provide automatic scoring and feedback. Complementary, essay and open-ended questions allow for the assessment of discourse-based subjects and higher order skills, and encourage productive and creative thinking. Following a mixed approach, teachers will not be forced to invest time and effort in rebuilding their personal question-banks while they will have the opportunity to have a progressive introduction to objective testing.

In addition, there is still the need for human markers, who are able to diagnose students’ learning needs and provide detailed feedback conducive to learning. CBA should take advantage of affordances that permit teachers to annotate, grade and analyze written work and closed-type answers (Plimmer & Mason, 2004). Hand-written markings and recommendations can help teachers retain the plurality of scoring semiotics so as to express their personal feedback freely. It has been shown that both teachers and students feel that teacher-written feedback is an important part of the writing process (Cohen & Cavalcanti, 1990).

4. Methodology

We deployed the aforementioned design directions, for the development of a CBA application called ‘MyTest’. The participatory design methodology applied consists of two phases (“We!Design”, Triantafyllakos et al., 2008). During the first phase, multiple iterations of the same concise and highly structured collaborative design session are conducted with different students. These sessions help participants to (a) recall and elaborate upon their prior problems and needs concerning educational assessment, (b) prioritize their needs and work together in order to suggest ways of satisfying them in the form of exact sequences of user actions and system responses (c) design a low-tech prototype interface, which incorporates those sequences, using a whiteboard, coloured markers and Post-itTM notes. In the second phase, the designers systematically analyze and then integrate student products from the different sessions into a final application.

Although this method is developed for use by students, we modified it to place the teachers in the students’ position. Nine design sessions were held over a five-day period, four with students and five with teachers. In total, 15 students from the 9th and 10th grade of a Greek high school participated, as well as 16 teachers specialized in different subjects. Students and teachers selected were computer literate, as the methodology requires. The general objective of the sessions was the design of an assessment application that would run on Tablet PC platforms and which should enhance previous practices by using both closed-type and open questions. Two questionnaires were administered at the end of each session which examined students’ and teachers’ beliefs about the design process (satisfaction with their participation, satisfaction with the methodology’s outcomes, freedom of expression during the design process) and the perceived usefulness and ease of use of each session’s interface prototype (as in Triantafyllakos et al., 2008). The sessions were also video-recorded for qualitative analysis.

MyTest application was compared with an assessment application which was developed using the same participatory methodology with the involvement of forty Greek undergraduate students divided in ten design sessions. The comparison was based on student-specified needs and proposals and was conducted with the aim of identifying the requirement (or not) for tailor-made assessment applications. Three researchers analyzed independently the functions of the two assessment applications and commonly-accepted dissimilarities were documented and interpreted.

In order to evaluate the MyTest application and the Tablet PCs’ adequacy under real learning conditions, a pilot evaluation with six assessment sessions and 67 students was conducted in a private secondary school. The duration of the evaluation was 35 min, the domain of the assessment was mathematics and the tests included twenty closed-type questions. Three sessions (31 students) were realized in a
typical classroom with the use of Tablet PCs. The other three sessions were run in a laboratory setting using typical computers. When students completed the test, they were asked to complete a short questionnaire (12 questions) concerning perceived usefulness and ease of use of the application MyTest and compare the assessment experience with the traditional paper-based assessments. All questionnaires included 5-point Likert scales.

5. Results

5.1. The ‘MyTest’ application

The average duration of the sessions was about two hours. Seventy-six needs were proposed by the students which were afterwards grouped and rephrased into a list of 19 needs and task sequences, based on the number of sessions in which they appeared and the participants’ average evaluation of their importance. A total of 91 needs were proposed by the teachers which accordingly were transformed into a list of 27 needs and task sequences. The two lists were merged into a common final one which included 25 assessment needs. The sessions’ products and prototypes can be found at the URL address: http://ierg.csd.auth.gr/We!Design. The assessment application MyTest was developed using DHTML and AJAX technologies while the components for handling “digital ink” were developed using C# and the Microsoft Tablet PC SDK. The application operates both on Tablet and typical PCs. Here, we will focus only on the main functional screen of the application (Fig. 1).

On the left-hand side of the screen we find the “question list” (I1) where a question can be in one of three possible states: (a) “unanswered” (b) “answered” and (c) “uncertain”. The student is able to move freely among the questions in a non-sequential way by tapping (clicking) a question on the list or using the “next” and “previous question” buttons at the bottom of the screen. Two panels provide time (I2) and progress (I3) information and can be hidden. The full question text is presented at the top, along with optional information provided by the assessment author (such as question difficulty and weight/score). Below the question we find the response area and the scrapbook (I4). In these areas the student can write anything using either the Tablet PC stylus or the keyboard. The student is able to customize his handwriting by setting the size and the colour of the digital ink. The background in the response area may also be configured in various paper styles, such as blank, lined or grid paper, in order to facilitate the handwriting of special responses. The handwritten text from the scrapbook can also be transferred to the main response area, with a drag ‘n drop move of the stylus. In the case of formative assessments, students may choose to submit the contents of the scrapbook along with the test to the instructor, as it can provide a clear window to their thinking and working strategies. When students are not sure about an answer, they can characterize it as “Uncertain” (I5). On the right-hand side of the screen there is a panel (I6) for supporting tools that the teacher may choose to provide for general use (such as a dictionary, a calculator, or a drawing toolbar for geometric shapes) or specific to a question (like instructions, hints, theory diagrams, or links to external resources). Both the “question list” and the “supporting tools” are collapsible so that students can have the greatest possible response area at their disposal when reading or responding. The Tablet PC offers the opportunity to choose between a landscape and a portrait orientation for the screen. The former provides a wider writing space, while the latter resembles more the look and feel of writing in a notebook and minimizes scrolling. In Fig. 2A we can see the portrait configuration of the application.

![Fig. 1. Main functional screen of the MyTest application.](image-url)
Objective tests are graded automatically and the student may immediately see the score and the feedback. In the case of essay questions, or whenever the teacher wants to provide manual feedback, a manual procedure is followed. Teachers can highlight and provide comments on any part of a student’s response using digital ink markers. Once the marking process is complete, the graded assessment is released to the student. Fig. 2B presents the grade awarded for an answer along with the teacher’s ink annotations on the side. Peer-assessment can also be supported, allowing students to grade their classmates’ answers, using the handbook or model answers provided by the teacher.

5.2. Evaluating the design experience

Participants’ evaluations of their experience in the design sessions and the corresponding interface prototypes are shown in Table 1 (variables computed according to Triantafyllakos et al., 2008). Independent t-tests did not reveal any differences between teachers’ and students’ evaluations. All participants indicated that the sessions were enjoyable (M = 4.90, SD = .30), interesting (M = 4.90, SD = .30) and felt free to express their thoughts without any kind of restriction (M = 4.90, SD = .30). Participants were satisfied with the application they co-designed with their colleagues (M = 4.45, SD = .80) and, most interestingly, they specified that they would rely more on educational software designed in this manner (M = 4.90, SD = .30). There was an apparent criticism of available educational software and concurrently an indication of their willingness to take over responsibility for their learning.

The participants were positive but a little bit more cautious about the usability and the utility of the prototype they developed. The reader should not forget that the low-tech prototypes were imperfect, sketch-based records of users’ proposals. They found their prototypes useful (M = 4.52, SD = .62) and specified that they would like to use them in the examinations of many courses (M = 4.61, SD = .62); however, they pinpointed an important concern: in most sessions, they did not feel sure about the difficulties that novice users would face (M = 3.23, SD = 1.43). They asserted that even a simple application might provoke problems in students’ interactions.

![Fig. 2. The MyTest application in portrait layout: (A) student’s (testing) view, (B) teacher’s (grading) view.](image-url)
The analysis of the video-recording revealed considerable differences between students' and teachers' concerns and suggestions and underlined the value of a balanced co-participation of the different stakeholders. Students felt comfortable and relaxed with the We!Design methodology from the beginning of the design sessions. They demonstrated a special focus in ensuring the readability of the questions and the answers and they discussed at length techniques for presenting detailed feedback on their answers and not just a mark or short comments. They indirectly disapproved of existing feedback practices and searched for essential means of transformation with the opportunity of the new mediation. They were creative and unpredictable, a fact that underlines the need for their influence e.g. "I want to write with T9, as I do with my cell phone", "I want to submit the questionnaire by writing 'the end' with the stylus", "the application should monitor the noise levels in the room and automatically produce remarks when they exceed some predetermined limits". Such needs were not incorporated in the application since they were evaluated negatively by the remaining participants in the session (a possible drawback of the We!Design methodology).

The teachers, initially, were more cautious and conservative trying to safeguard the assessment process and discussed issues such as cheating, shuffling the questions, disabling the activation of other software etc. They demonstrated a particular interest in assistance for students with special learning needs proposing solutions such as personal adaptation of the time available, recording of oral answers etc. When teachers from different domains participated in a session, it became more effective and efficient; different perspectives of the intrinsic characteristics of the assessment process were raised and had to be composed into tangible needs.

5.3. Comparing high-school with university students' needs

Undergraduate students were also upbeat about the design sessions while they had more complex expectations and concerns regarding the assessment process. Their application is shown in Fig. 3. Distinguishing features identified between high-school and university students' needs:

- **management of the assessment process (I1):** the undergraduates perceived more sophisticated schemes of time management (multiple alerts with sliding bells in the time bar, analytical presentation of time-response performance indicators during the test), emergency management (technical problem alerts, recovery strategies) and appearance management (design templates, layout configuration) processes. They focused on providing to the users more control over the assessment process and we can presume that those dissimilarities emanated from their more developed metacognitive learning skills.

- **complexity and information load on the interface (I2):** The undergraduate students designed a more adaptable layout (e.g. display of all the questions in one scrollable screen) which included impressive interaction objects (e.g. navigational scrollbar with semantic notation and colors corresponding to question status, multi-tab area with complex arrays for reviewing and navigating to questions of various status). They tried to provide diverse ways of viewing the questions and navigating among them. This can be attributed to their more extensive computer experience which allowed them to explore and exploit the interactivity language better.

![Fig. 3.](image-url) Another assessment application designed by undergraduate students using the We!Design methodology.
• **communication with instructors (I3):** the undergraduates incorporated communication facilities in the application in order to converse electronically with the teachers during the examinations when they have doubts about the questions (e.g. shared discussion board, private but publishable discussions, anonymous comments). It was obvious that they had confronted many difficult circumstances in the past and that they conceived open electronic discussions with teachers as a significant prerequisite for fair and calm participation in the assessment process.

Diversity could also be attributed to the particular context of the assessment within higher education and within Greece. Undergraduate students did not emphasize the learning aspects of the evaluations but concentrated on creating the optimum conditions for extracting the best performance; they aimed at achieving the highest score.

The way of addressing those differentiations is unique in our approach since they are revealed and formalized by the stakeholders themselves and not hypothesized by the designers. We also assume that there are multiple variations of those needs among different schools, academic departments, countries and cultures. The participatory methodology managed to reflect in the MyTest application differences in cognitive development, computer experience, assessment experience and priorities in those two educational environments. While the two applications may not be optimal since they could incorporate in both cases elements enhancing students’ experience, they should not, on the other hand, provoke any undesirable reactions which could challenge the transformation of practice. This comparison exemplifies how easy it is to make that mistake.

### 5.4. Pilot evaluation of the application ‘MyTest’

Students’ formative evaluation of the MyTest application is shown in Table 2. In general, students in both testing conditions seemed positive although significantly less enthusiastic than the students and teachers who participated in the design sessions. That was anticipated since they did not have the feeling of ownership that influenced the ‘designers’ evaluations while they also interacted with an existing application and not a prototype. Interestingly, the variables “utility” (including questions such as “I consider the application myTest as especially useful”, “I prefer tests that use the assessment application rather than the traditional way”) and “ease of use” (e.g. “The application is not easy to use”, “the interactions with the application are vague and obscure”) for the application ‘Mytest’ were perceived differently in the diverse platforms. Students evaluated the same application as more useful ($t(65) = 3364, p < 0.002$) and usable ($t(65) = 3368, p < 0.001$) in the Tablet PCs than in the typical PCs. This fact supports our claims about the suitability of the Tablet PC platform since it rendered the assessment application friendlier for students. The students, independently of the platform used, encountered with a relatively neutral to positive attitude the general comparison of the application with traditional paper-based exams (e.g. “Answering in MyTest is easier than in paper-based exams”, “Exams in MyTest are easier than paper-based exams”). That was our main objective from the beginning. The assessment method should not change much and probably would be accepted straightforwardly in its new form. We argue that it is exactly at this point that the transformation of the assessment practices with the incorporation of new modes of assessment should start.

### 6. Discussion and implications

Three main guidelines were highlighted for the incorporation of CBA in secondary education: design in cooperation with the students and teachers, selection of the most adequate media platform and planning an evolution rather than a revolution of the prior practices. We provided evidence that designing effective CBA applications can be realized by actively involving students and teachers in the design process. Both students and teachers were excited about their participation in the design sessions, and they asserted that they would rely more on educational software designed using this approach. This throws light on their willingness to undertake a process of re-conceptualizing existing pedagogies in the light of new opportunities and engage in the co-formation of their future. Such approaches support a decentralized future that empowers locality and diversity, participation and attitude and comes in opposition to homogenization and impassivity. Certainly, those approaches also conceal dangers since they can invoke simplification, disguise potential, and may be validated only because of their phenomenal “political correctness” and not assessed in pure educational terms.

The application presented employs flexibility at multiple levels; as it can be used for different types of tests (from informal self-assessments to credit-bearing summative assessments), on different computer platforms (from simple to “digital ink”-enabled PCs), can support all common question types, providing various levels of scoring and feedback (automated, teacher-based or both). It retains the essential advantages of CBA such as saving staff time in marking and feedback for at least the closed-type questions, maintaining history of students’ results and teachers’ comments, allowing the development of reusable question banks etc. The Tablet PCs seemed to respond better to the advantages of CBA such as saving staff time in marking and feedback for at least the closed-type questions, maintaining history of students’ results. It is important to note that the proposed design principles may not always drive the development of innovative assessment applications but can at least ensure that they will correspond in an efficient and effective manner to the students’ and teachers’ skills and expectations.

---

**Table 2**

<table>
<thead>
<tr>
<th></th>
<th>Total</th>
<th>Tablet PCs</th>
<th>PCs</th>
<th>$t$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td><strong>Utility</strong></td>
<td>3.93 (.97)</td>
<td>4.30 (.46)</td>
<td>3.60 (1.16)</td>
<td>3.364</td>
<td>&lt;0.002</td>
</tr>
<tr>
<td><strong>Ease of Use</strong></td>
<td>3.36 (.87)</td>
<td>3.71 (.61)</td>
<td>3.06 (.94)</td>
<td>3.368</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Compare CBA-TA</strong></td>
<td>3.42 (.89)</td>
<td>3.48 (.74)</td>
<td>3.36 (1.00)</td>
<td>-.561</td>
<td>.577</td>
</tr>
</tbody>
</table>

*Correlation is significant at the 0.01 level at (two-tailed).*
thus facilitating the smooth integration of CBA into the reality of secondary schools. The applications’ success is greatly related to the technological, social and cultural characteristics of the environment in which they are designed. Certainly, pedagogical, technical, operational and financial feasibility must be further analyzed, not only for the strategy in the short term, where a demonstration of a pilot is easier but also in the medium to long term, if expansion is to take place.

It should also be noted that students are not experts; they have limited design skills and are incapable of knowing, experiencing and taking advantage of pedagogical approaches and pioneer technologies. Their contribution and their assessments, are guided by experience and intuition and can be simplistic and, at times, unoriginal. Following that, our evaluation of the assessment application should be complemented using empirical measures, other than the self-reporting questionnaires that were used to collect student and teacher opinions.

Our efforts will concentrate on extending the installation base of the application in schools that have or are willing to acquire Tablet PCs and in evaluating more analytically the many dimensions of students’ assessment experience, such as attitudes after repetitive use, platform and test-mode effects of Tablet PCs, the effect of digital ink writing technology on open-ended answers etc.

References


Muter, P. (1996). Interface design and optimization of reading of continuous text. In H. van Oostendorp & S. deMul (Eds.), Educating the electronic generation: Technological, social and cultural characteristics of the environment in which they are designed. Certainly, pedagogical, technical, operational and financial feasibility must be further analyzed, not only for the strategy in the short term, where a demonstration of a pilot is easier but also in the medium to long term, if expansion is to take place.


