

## Mobile Phones for Spain's University Entrance Examination Language Test

Jesús García Laborda<sup>1\*</sup>, Teresa Magal Royo<sup>2</sup>, Mary Frances Litzler<sup>1</sup> and José Luis Giménez López<sup>2</sup>

<sup>1</sup>Departamento de Filología Moderna, Universidad de Alcalá, calle Trinidad, 3, 28801 Alcalá de Henares (Madrid), Spain // <sup>2</sup>Departamento de Ingeniería Gráfica, Universidad Politécnica de Valencia, Camino de Vera, s/n, 46022 Valencia, Spain // [jesus.garcialaborda@uah.es](mailto:jesus.garcialaborda@uah.es) // [tmagal@degi.upv.es](mailto:tmagal@degi.upv.es) // [mf.english.uah@gmail.com](mailto:mf.english.uah@gmail.com) // [jojilo@degi.upv.es](mailto:jojilo@degi.upv.es)

\*Corresponding author

### ABSTRACT

Few tests were delivered using mobile phones a few years ago, but the flexibility and capability of these devices make them valuable tools even for high stakes testing. This paper addresses research done through the PAULEX (2007-2010) and OPENPAU (2012-2014) research projects at the Universidad Politécnica de Valencia and Universidad de Alcalá (Spain) to provide a powerful but low cost delivery system for the foreign language paper of the Spanish College Entrance Examination (henceforth PAU). The first project, PAULEX, intended to create a robust mobile platform for language testing while the second, OPENPAU, examined the specific applications of ubiquitous devices to create more dynamic forms of assessment. This paper focuses on the projects' design, testing theory, and technical evolution including visual ergonomics. The current results demonstrate the technical and didactic feasibility of mobile-based formal assessment that aligns student needs with the kind of inferences that the mobile based language test should provide academic authorities.

### Keywords

Mobile learning, High-stakes testing, College entrance examination, Foreign language, Higher education

### Introduction

Mobile phones have been playing an increasingly significant role in education in the last years, and although until recently very few tests were delivered through them, their flexibility and capability to do so have suggested their potential even for high stakes-testing. High-stakes testing can be defined as those tests with important consequences for the test taker such as acceptance to university, a scholarship, or a license to practice a profession, all of which may have a great influence of the testee's life. Using mobiles beyond their traditional uses such as podcasts, mp3 applications, and even learning apps seems to be a real challenge at this point, yet they have already been used for language testing as in PhonePass, previously called SET- 10 ([http://www.7act.net/7ACT\\_files/set10.pdf](http://www.7act.net/7ACT_files/set10.pdf)) test. The validity of the test has repeatedly been supported (Downey, Farhady, Present-Thomas, Suzuki & Van Moere, 2008) but little evidence has been provided of its operativeness in real educational contexts. As a consequence, the potential opportunities for mobiles for language testing are still open (Valk, Rashid, & Elder, 2010) but sound projects need to be implemented. This is the kind of research that the Ministry of Education started to support in 2007. By that year, the Spanish and regional educational authorities responsible for the high stakes University Entrance Examination ("Prueba de Acceso a la Universidad", henceforth, PAU) had determined the need to design a new test with greater validity than the current paper-based test, which only included the traditional tasks of reading and writing along with grammar questions. The new test had to include listening and speaking activities. However, budget cuts reduced the chances of implementing a new exam that could include speaking and listening activities unless a low-cost possibility could be found. With a view toward designing a modern test in the hopes of saving the possibility of including the two skills, the Universidad Politécnica de Valencia obtained funds for the development of an online testing system (PAULEX, "PAU en Lenguas Extranjeras") project from 2007-2010 and the OPENPAU ("PAU abierta") later on between 2012-2014 whose results were described by García Laborda (2012). After testing over 150 students online it was concluded that implementation of the computer test would save human resources and be economically feasible in a period of two to three years. As a subproject, the PAULEX project addressed the use of mobile phones (García Laborda & Giménez López, 2010), which is the main focus of this paper.

### Literature review and theoretical approach

When addressing this sub-project, the research team felt that the use of mobiles phones, like any other delivery system, could not challenge three main testing features: validity, reliability, and practicality (Bachman & Palmer, 1996, among others). Validity here means that if a student gets a score of X on the test, it means that he should be

able to study using that foreign language at university; reliability provides information on the “the precision of the test measurement” (Salmani-Nodoushan, 2009, p. 1); and practicality implies that the test can be implemented in real life. In addition, the test construct in which mobiles are to be applied has taken into account current theories in language testing and Communicative Competence (Canale & Swain, 1980; Canale, 1984; Bachman & Palmer, 1996). Thus, the questions considered were (1) why use mobile phones for language testing? (2) How can the basic testing features be assured? (3) What learning theories are implicated in their use?

For most part, the collection of evidence in both projects was based on Weir’s validation framework and the Evidence Centered Design (Mislevy, Steinberg, & Almond, 2002; Mislevy & Haertel, 2006). Weir (2005) feels that the reliability of a test depends mostly on its conditions for validation. For him, it is necessary to have warrants that there will be two main types of validity in implementing a test: context validity and theory-based validity. Both are interrelated and need to be considered interdependently. Context validity is divided into three parts: task, setting and administration, and task demand (similar to test construct), while theory-based validity includes executive processes and executive resources. Given this framework, the theoretical application for the mobile application of the PAU took into account the aspects included in figure 1

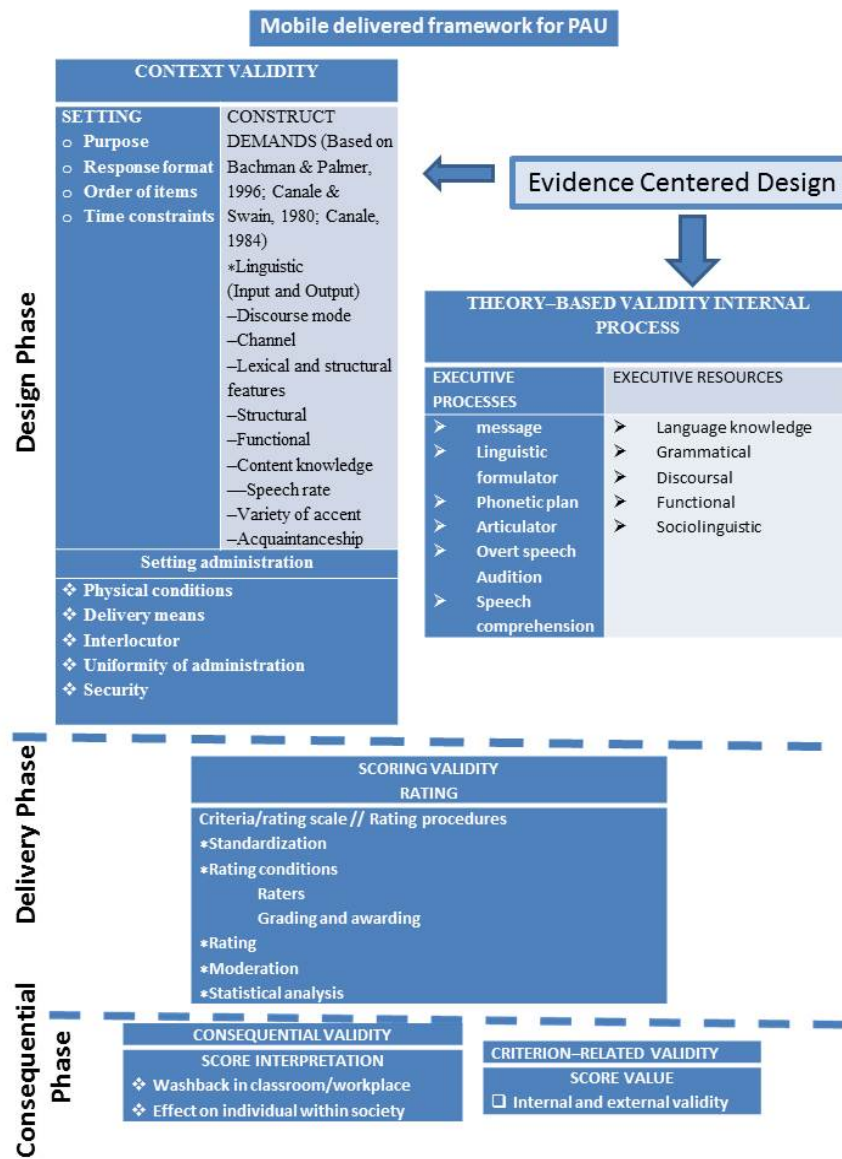


Figure 1. Delivery framework (based on Weir, 2005)

As indicated in figure 1, the theoretical aspects that needed to be considered in the implementation of the mobile-based test were grouped into three main phases: design, delivery, and consequential. The limitations and scope of this paper only allow for a discussion of the most important aspects of the design phase. Within this framework, this paper will mostly focus on contextual validity because this is where Weir places the delivery system factor. However, there is no question that the implications of using certain delivery systems -whether pen-and-paper or mobiles- are present in all the aspects presented in figure 1. This approach is based on performance-oriented tasks, which intend to resemble communicative acts of the language. In designing the test process in the test, two main options were included: a cognitivist and a social constructivist intervention. While the cognitivist approach supports the notion that students bring knowledge at the time of testing and this is represented by observable behaviors, the social constructivist approaches in language testing are very much related to the development of the Zone of Proximal Development (ZPD) (for further discussion see Poehner, 2008) through the examiner's intervention and moderation. The Zone of Proximal Development is defined as "...the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers" (Vygotsky, 1978, p. 86). Based on the development of the ZPD, Poehner (2008) also suggests that Dynamic Assessment includes both approaches. Both PAULEX and OPENPAU projects follow the principles of Dynamic Assessment (DA). This approach considers that students have a current level of language knowledge (the one they show without test moderation) and a potential level (the one that they show when their output is computer or human-based moderated and improved through this interaction). According to DA, both can be included in a two-part assessment if the first is moderated (socio cognitive approach) and the second just serves to obtain current language evidence (cognitivist approach) without any tester's intervention. This can be seen the process shown on figure 2.

According to these principles and given their experimental nature, the PAULEX and OPENPAU projects placed more emphasis on achieving a sound design based on experimental evidence than on potential achievement scores through the use of the test. In practice, evidence was collected and recorded through the use of mobiles. There were five main benefits that justified the decision: (1) the lower cost of mobile based hardware; (2) immediacy of rating and results; (3) ease of recording during oral interviews (hence, data available for further revision of the test and research); (4) the candidates' familiarity with the delivery means; (5) possibilities for students to rehearse; and (6) ease of rating and administration. Additionally, accessibility for schools and/or official testing centers would enable the optimization of space.

Evidence obtained from the tests, which was moderated, was processed through the Evidence Centered Design (figure 1) (Mislevy, Steinberg, & Almond, 2000; Mislevy & Haertel, 2006) after the first interview, and implied the design of adequate tasks that considered all the linguistic requirements (as seen in figure 1) and could be delivered through mobile phones. According to the cognitivist approach, used in the second testing session, tasks had to be automatically delivered and recorded without moderation to provide current real data. Then the responses were rated (and the scores validated) and with a view toward having an impact on decisions for teaching and high stakes decisions.

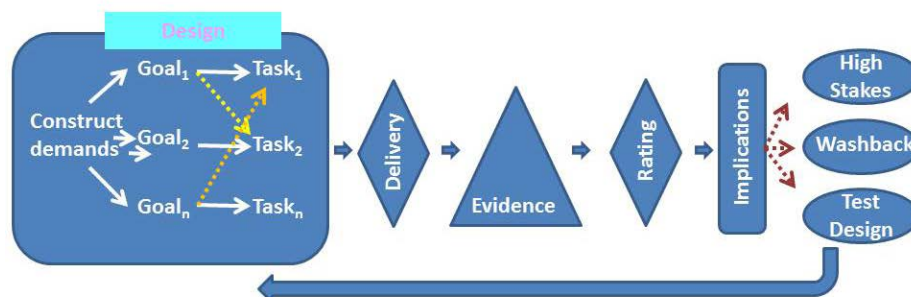


Figure 2. Development of test process

Overall, the research team felt that the use of mobile phones was strongly founded but they recognized that the advantages and disadvantages needed to be weighed. Table 1 presents the pros and cons of their use:

Table 1. Use of mobile phones according to test characteristics

Test characteristics	Advantages	Disadvantages
Place and time circumstances	Convenience of location and time because they require no external human presence	Security and technical assurance of full functionality required; otherwise test is at stake
Test rubric and process for responses	Tasks are consistent and delivered exactly in the same way to all tessees	Tessees may have different levels of understanding of the instructions according to proficiency levels
Test delivery	Current developments in mobile phones increasingly enable the inclusion of audio and image, thus enhancing contextualization and richness of responses	Standardization of mobile phones used for the test is needed; otherwise significant differences in responses can be found even by the same speaker.
Construct, rating, and scoring	Automatized rating validates the equanimity in multiple choice items; separation of the rater from the testee enables rating protocols to be followed without the influence of contact with the testee. Thus assessments are more objective.	Human protocols do not assure complete equanimity (Baldwin, Fowles, & Livingston, 2008).

### Mobile phones in high stakes testing: The PAULEX project (2007-2010)

Since the implementation of the originally planned computer-based language testing platform was costly -albeit assumable in the long term-, one of the suggestions for the researchers was the use of mobile phones for the Speaking test only until the online platform could be used. However, while mobiles were originally thought to support student training, almost from the beginning the project management felt that they could also have a very positive effect on learning and they could encourage after-test washback effects. The main reasons to implement mobile phones were that the hardware was less costly than for computers, their use could be more accessible as they can easily be delivered and collected to and from each school, and their use could facilitate rapid assessments by testing units (which would resemble calling centers in their functioning and organization). These testing units could potentially organize and deliver a large number of tests in a limited time. The tests could be delivered automatically; the students' responses could be recorded and assessed later by human raters.

From the beginning of this three-year project, it was clear that a well-trimmed double design project was needed for the delivery, ergonomics, and content inclusion. Figure 1 describes the organization of the PAULEX project.

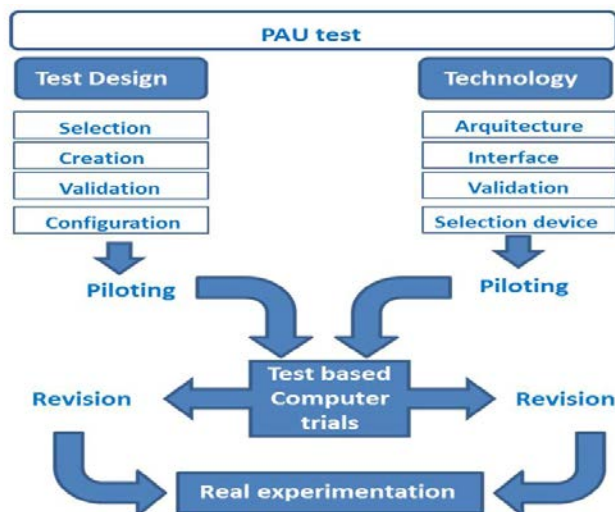


Figure 2. Organization diagram of PAULEX project

As can be seen in figure 2, two branches were organized: one devoted to the linguistic and validation aspects, and the other focused on the technological design of the online and mobile platforms. From the beginning it was clear that most of the significant difficulties were associated with the test design since the technology group had already been involved in similar projects before. Because the validation process was central to the project, the mobile application was designed and tested considering a variety of students and also bearing in mind that the PAU project served to obtain inferences of whether students would be able to use English for university work. Furthermore, the mobile technology branch considered that not all students have the same ability in using mobile technology so the technological specifications were relevant and accessible to students with special needs.

The development of the mobile phone subproject within the larger PLEVALEX project was intended to provide information on three aspects: (1) student adaptability to the new environment, (2) content and test validity for the listening/speaking tasks, and (3) delivery reliability. As mentioned above, mobile phones have been thought to foster learning more than to be used to assess students. Learning would take place by providing them with test samples that could be used anywhere and at any time. In this way mobile phones would bring to the fore the required testing skills in combination with similar listening and speaking tasks along with affective considerations but in a more interactive and usable manner. This process would also provide opportunities for authentic learning and the elimination of test fear would probably favor motivation. In this sense, the mobile phone sub-project sought to engage students in terms of motivation, high stakes test practice, and language learning. The results for this project were obtained through triangulating linguistic achievements, field notes, and a usability analysis carried out through a 20-item questionnaire computer delivered to all the students who took part in the research (García Laborda, Giménez López, & Magal Royo, 2011).

### Validation method

We used five types of validation analysis for the PAULEX project. First, we did a Delphi analysis (Custer, Scarella, & Stewart, 1999) to foresee potential issues in the mobile phone test with experts and then a reduced number of regular users (3). Second, we observed the intended scores of the mobile phone users and compared them with those obtained with the online platform (Sariola, 2003). Next, we analyzed the video recordings from the pilot studies. After that, inspection techniques were followed to do a usability analysis (Nielsen & Mack, 1994). Finally, the students' attitudes toward mobile phone use for the test were analyzed (as seen above).

### Usability analysis

As discussed above, the first test was conducted as a pilot test with a small sample of students, to detect faults in the design of the application and to debug and test the viability. After making some corrections such as adapting some aspects of the content and navigability, a second review of the application was made using potential users. In this second test the number of the sample was expanded to 144 individuals in the last year of high school (aged from 17-18 years), all of whom lived and studied in the area of La Oliva-Gandia (Valencia, Spain) (see table 2).

*Table 2. Students' school of origin*

	High school	Frequency	Valid percentage
1	IES Tirant lo Blanc	20	13,9
2	IES Monduver	22	15,3
3	IES Veles e Vents ,	33	22,9
4	IES Maria Enriquez	29	20,1
5	IES Ausias March	40	27,8
	Total	144	100,0

To evaluate the usability of the interface, a likert scale questionnaire ranging from 1 to 4 (to avoid indecisions) was used.

## Results of the second test

Once collected, the data were processed using the SPSS statistical program. The first part of the test, which related to knowledge of the environment, focused on aspects to justify routine use and availability of phones, adaptation to the environment of the test items, and utility-satisfaction.

*Table 3. Students' attitudes toward the mobile-based tool operability*

Responses	Frequency	Percentage	Valid Percentage
Totally disagree	9	6,3	6,3
Disagree	35	24,3	24,3
Agree	89	61,8	61,8
Totally Agree	11	7,6	7,6
Total	144	100,0	100,0

*Table 4. Students' attitudes toward the mobile-based tool usefulness*

Responses	Frequency	Percentage
Totally disagree	9	6,3
Disagree	16	11,1
Agree	100	69,4
Totally Agree	19	13,2
Total	144	100,0

*Table 5. Students' attitudes toward the mobile-based tool time facilitator*

Responses	Frequency	Percentage
Totally disagree	10	6,9
Disagree	21	14,6
Agree	69	47,9
Totally Agree	44	30,6
Total	144	100,0

Once the descriptive data had been surveyed and the group statistics had been examined, it was determined that the results were satisfactory as a whole. The results are above 1.5 on average (on a 0-3 scale).

*Table 6. Students' attitudes toward other factors in relation to mobile phone use for language testing*

Responses	Valid	Mean	Standard Deviation	Mode	Min.	Max.
A mobile-based task design helps me to perform better	144	1,71	0,058	2	0	3
A mobile-delivered test is useful	144	1,90	0,058	2	0	3
Mobiles help to save time in taking this test	144	2,02	0,071	2	0	3
Mobiles are adequate to cope with my needs for this test	144	1,74	0,060	2	0	3
I learned to use the application quickly	144	2,44	0,057	3	0	3
It is easy to remember how to use the application	144	2,45	0,053	3	0	3
I became familiar with the application easily	144	2,32	0,051	2	0	3
I think this is a good application	144	1,94	0,062	2	0	3
It is user friendly	144	1,94	0,060	2	0	3
The application works as I expected	144	1,92	0,059	2	0	3
I would recommend its use to other students	144	1,90	0,075	2	0	3

As observed, in general the students valued the use of mobile phones for language testing very positively. The results of these tests have led us to continue with the research, which is still currently being developed.

### **Advancing toward solutions for the PAU: The OPENPAU project (2012-2014)**

In the years between the PAULEX project and the beginning of the OPENPAU project, Spain started experiencing one of the worst financial crises in its history. In that context, the research team of the PAULEX project observed that mobile phones would be a valuable asset in testing oral skills (speaking and listening) and reading efficiently at a low cost. However, the team also considered that mobiles would be inappropriate for writing due to the intrinsic difficulty of keyboard use (García Laborda, Giménez López, & Magal Royo, 2011; Park, 2011). The design principles for developing a mobile phone application were the following (also Keskin & Metcalf (2011):

- Use of video communication with the examiner or a video delivery system if videos are used (possibly the most likely situation),
- Creation of a podcast library for student test preparation,
- Adequate real or deployed time access and adequate connectivity,
- Augmented reality possibilities,
- Mobile Blackboard or a similar platform for test preparation.

The first trials on a large scale are expected to begin by September 2013. The technology is designed to incorporate these conditions. The following sections address this concern. Thus far only the Delphi analysis and a very small sample of research have been complete.

### **Significant results of the PAULEX and OPENPAU project for m-testing technology**

The results hereby presented are mostly related to the observations and research undertaken between the end of the PAULEX project and the beginning of the OPENPAU project. However, the ideas are based on the results from PAULEX and the triangulation of the Delphi method and focus groups. From their reactions and opinions, we concluded that the different kinds of interfaces for mobile devices favor tasks such as speaking and listening, even multiple choice tasks but are rather limited for reading and especially for writing. In the case of online exams with different kinds of tasks, adjustments must be made to navigation and content so that users can feel more comfortable when viewing and inputting information or data. The adaptation of user interfaces for certain tasks in a limited period of time requires prior understanding of certain determining factors such as the physical, functional, and formal accessibility of the application. For example, an interface with a hierarchical menu on a mobile phone is useful for beginning users because the appropriate options can be selected through the presentation of a series of menus. Hierarchical menus require relatively higher numbers of key clicks but this is acceptable for novel users who need help using unfamiliar navigation systems and thus leads to diminishing differences due to technology knowledge and serves to validate the use of mobiles as delivery system (Weir, 2005).

### **Interface design**

The most recent interfaces developed have been designed with specific criteria for users taking the university entrance exam. The fundamental criteria studied in this period were accessibility, ergonomics, and the functionality or usability of the application.

#### *Accessibility*

Both projects followed the criteria for technical accessibility for interface design proposed by the World Wide Web Consortium, W3C, so that they reached the largest possible number possible of students as end users including those having visual or auditory impairments. This was achieved mainly by following the Web Content Accessibility Guidelines, WCAG 2.0 (<http://www.w3.org/TR/WCAG20>). In terms of accessibility of the contents of the university entrance exam, we considered the type of programming language used for navigation and also established guidelines for information access (Nelly et al., 2009). The first applications developed, PAULEX and OPENPAU, were created

in environments accessible on Internet with access to the contents delivered online by way of contextual menus for accessing the task management area and student management area. As for the exams created for the students, special attention was paid to visual and functional accessibility of navigation during the final exam tasks. The tests permitted us to determine the potential for mobile phones and the testees' acceptance mobil phones (Magal-Royo, Fajarnes, Tortajada Montañana, & Defez Garcia, 2007). At the same time, they revealed the importance of two determining factors in the development of future applications: the present rate of technological progress of ubiquitous devices, and the adaptation of contents to the restrictive conditions imposed by them.

### *Ergonomics*

The ergonomic aspects examined for the different applications created for mobile devices focused on the visual ergonomics that enabled students to focus easily and effectively on completing the different tasks. Various studies conducted after the experts' research revealed the need to establish formal visual guidelines for the content of language learning tasks to enable navigation that is directed and transparent (Weining, Heng, & Guoping, 2007; García Laborda, Magal-Royo, de Siqueira Rocha, & Álvarez, 2010). Ubiquitous devices (mobiles, PDAs, smart phones, netbooks, etc.) can be small in size. That is, they have small screens that limit the space for user interface and the information available on them thus writing and reading tasks may have an additional difficulty due to the fact that a global vision of the read or written text is always desirable. In fact, the information shown must be carefully selected and presented so that it facilitates user interaction not only with the device but mostly with the task content. The major problem with the large variety of screens on ubiquitous devices is the direct impact on information access and visualization because no normalized standards have been established so far (Chae & Kim, 2004; Piolat, Roussey, & Thunin, 1997). This problem was of the most significant ones in test validity. In theory, if the test is implemented, the Ministry should provide all the testees with the same mobile phone to avoid biased or unfair testing conditions.

### *Limited data input mechanisms*

These devices are also limited in terms of data entry procedures because of their reduced size. The methods used most often for mobile devices nowadays are the keypad, which has more than one function associated to each key, and the touch screen. Both methods require a high degree of attention on the part of users and can lead to errors, a situation that limits how they can be used.

Thanks to improvements to user interaction mechanisms now found on ubiquitous devices, different channels can be used for data input, which can be simultaneous, synchronized, or combined for certain tasks. In the case of the mobile phones, which have different kinds of exercises (oral, comprehension, writing, etc.), data entry mechanisms can slow down or directly affect completion of the exam, for example, on the reading test, which can involve reading a long text or typing using a virtual physical (Giménez López, Magal Royo, García Laborda, Garde Calvo, & Prefasi Gomar, 2009) (see figure 3). The approach to design this interfaces was taken from the socio-constructivist theory of language that uses images to trigger the testee's output and the visuals to support and enrich the production. This can be also the case when the videoclips are interrelated in semi-interactive conversation through short questions or even in connection to other user to make dialogues between two testees in which the potential knowledge is visible after reconstructing one's production (Poehner, 2008). At the same time, a robust recording system and clear rubrics support the cognitivist approach that can be best seen in the long responses for descriptions or the multiple choice responses in which the students need to show evidence of knowledge without external mediation or support (what has been called current knowledge).

### *Usability*

In terms of the usability and functioning of the applications adapted to mobile phones created for the PAULEX project, the results show that the students considered it to be useful because it enabled them to save time while taking an exam of this kind (see figure 3). It was also determined that the students learned to use the application for mobile phones faster and independently due to their familiarity through daily use of mobiles, which enabled them to adjust quickly to it and to its guided interactivity. Analysis of the data related to level of satisfaction with the use of the



application was very high when there was a sensation of predictability that leads to a fast understanding of the method and learning how to do specific tasks on the mobile phone.



Figure 3. PAULEX application on mobiles

The overall conclusion of these first trials was that the students felt comfortable with the format (bearing in mind the limitations of the devices). The oral tasks with video presentations were evaluated with the same degree of confidence and reliability as the analogous activities on the web platform for personal computers.



Figure 4. Usability test of the PAULEX application

### Proposals for the design of an m-testing platform

The proposals in this section were also applied to the OPENPAU project and any future project of a language testing m-platform and are strongly based on the findings from the PAULEX project. The OPENPAU project has incorporated the application to the HTC Desire model mobile device whose base technology allows multimodal use of different forms of data input and output (see figure 3). To do so, a study was carried out in advance to determine the initial conditions needed for completion of the tasks on an English language skills exam. These included the

functional structure and verification of the sections in order to ensure efficient, real navigation. Afterwards, the following premises were considered vital for the m-application:

- Exam access should be possible using the keypad and/or the touch function for navigation on the mobile device, enabling access to each of the parts of the exam and including the possibility of repeating tasks as long as the user has not finalized the section.
- There should be a final screen for confirming that tasks that have been completed but it does not have to indicate whether they have been finalized by the user.
- The user should be able to select in advance the method of interaction or main form of navigation to change screens and tasks and to finalize them.
- The user should be able to see and hear any digital content annexed to the tasks created using the device's generic speaker.
- The user should be able to see and hear any digital content annexed to the tasks created using the external earphones and microphones.

The general format of the application contained the following visual sections:

- The program header area. The application name and official logo of the program participants appear in this section of the screen.
- The user data area. This area is fundamental for the final coding of the exam and student for initial correction and any future corrections, as well as any official reviews required by law at the national level.
- Area for viewing progression through the exam. This area has numbers indicating the different tasks that must be completed on the exam. This section will enable the students to know their progression throughout the exam from the point at which they enter their application access code until they send the completed exam. It starts with the reading of the student's data before the actual completion of the exam and provides information throughout completion of the exam including selection of the interaction mode, and completion of the different tasks on the exam, etc.
- Test area. This area shows the questions or exercises to be completed on each of the tasks. The content will vary depending on the functional and/or content characteristics of the exercises.
- Help area. This section will show general as well as specific information about how to complete the exam including the maximum score assigned to each section.

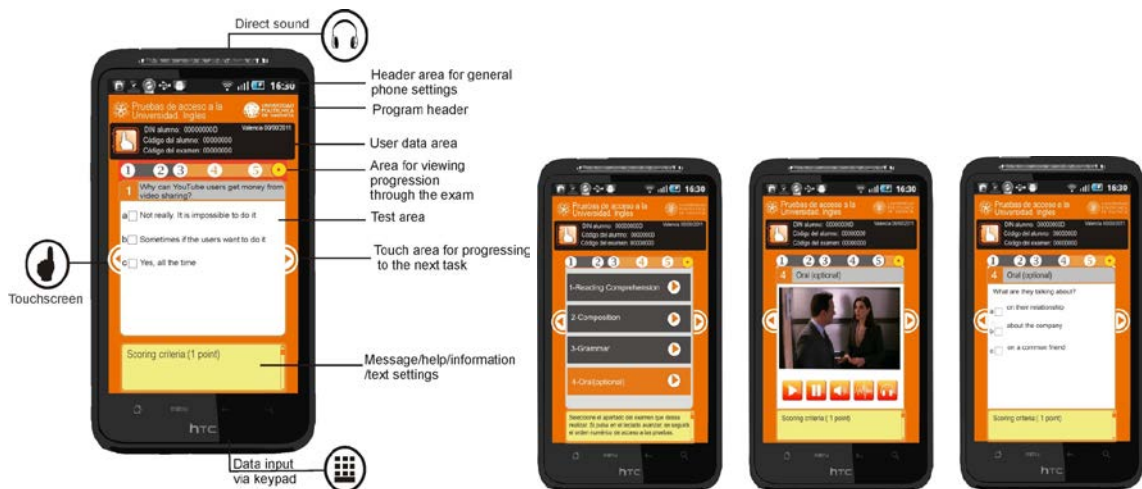


Figure 5. PAULEX application for mobiles, 2012

## Project results and conclusions

As observed in the PAULEX project, in situations of high stakes tests with a large number of students, mobiles have some advantages that may put them ahead of other testing systems in terms of budget, accessibility, familiarity, and sound quality. Additionally, although the results in the PAULEX project were limited, the validation methods provided information about the ergonomics, usability, integration, and motivation of the application. According to the data obtained, it was observed that prospective research should include the following aspects:

- Task adaptation to new types of mobile phones;
- Multiplatform systems;
- User satisfaction;
- External validity as compared to other delivery systems and other tests including similar pen and paper versions;
- Technical advances in software design;
- Pedagogical benefits;
- Delivery reliability;
- Functionality.

The students were eager to use mobile phones for language teaching and learning, but they mostly wanted to use them for speaking and listening. Still, the multiple choice items for grammar were also well regarded. However, the students predictably indicated that reading and writing were too difficult to be implemented, with reading rated in a better position than writing (García Laborda, Giménez López, & Magal-Royo, 2011). The PAULEX project also showed that mobiles were excellent for test preparation and an even more encouraging finding is that they offer great opportunities for the real test itself because the students would accept using them for real testing tasks. All three teachers indirectly involved in piloting their use supported mobiles and liked the sequencing and delivery procedure for questions, but they claimed that they had no software up to that time to implement the teaching at a large scale. They also found that, although the testing system could, in fact, be valuable to assess oral skills in the PAU in the long term, phones with bigger screens were desirable. At the same time, they doubted that the Ministry of Education would spend large sums of money on the terminals. However, they believed that the listening and speaking sections could be done online while the rest of the test could be done with pen and paper in order to lower the cost. Additionally, they mentioned that one set of mobile phones could serve more than one high school and maybe more than one year given adequate hygienic measures. Finally, they mentioned the convenience for raters since they could work from a distance either on synchronous or asynchronous testing.

Our experience also determined that technologies for developing user interfaces should focus on the requirement to offer simple interaction modes that are highly natural and adapted to future terminals and communication networks (Oviatt & Cohen, 2000). It is in this area in which technologies face their biggest challenge: attempting to integrate different modes of communication (visual, oral, auditory, gestural, etc.) in order to offer new more powerful methods of interaction with the user, grouped under the name of natural or multimodal interaction, thus overcoming the limitations of interfaces available today (Oviatt, 1999). The ultimate objective of natural interaction is to enable users to be able to use all the communication resources available to them, combining multiple modes of interaction and, therefore, creating a multimodal environment for information access (voice, audio, graphics, video, keypad, electronic pencil, pointer, mouse, etc.) (Oviatt & Larson, 2003). In this sense, the OPENPAU project is currently being driven by practical concerns. The current research is now exploring the potential for implementation and the pedagogical implications while extending the domains of the project to make it a multiplatform one. The study has shown the feasibility of using mobiles for the intended purposes and that the cost could probably be lower than the traditional face-to-face interviews while also permitting a better distribution of space for delivery and adequate rationalization of testing times. Most of the students might also engage in this testing means more easily than in a face-to-face interaction with the examiner. With the development of the OPENPAU application for ubiquitous devices, it has been found that technology has now progressed sufficiently to propose the offering of exams using multimodal access. The incorporation of new modes of interaction such as voice recognition for navigation, the use of touch screens, or synchronized use of the keypad will enable users more comfortable access in accordance with their needs and, thus, solve problems related to accessibility to the media (Magal-Royo, Giménez-López, Paíry, García Laborda, & Gonzalez-Del Rio, 2011; Magal-Royo & Giménez López, 2012).

Progress in the use and research of mobile phones for language learning is receiving increased attention and their use in Mobile Assisted Language Learning (MALL) is an area of steady growth. Despite the advantages this area offers users in terms of the flexibility and ubiquitous nature of the device and environment, as well as advances in mobile applications and Internet access, it must still deal with the need to seek efficient adequate interfaces for user needs for information access and transfer. In the specific case of task completion or specific processes, it is important to evaluate the impact of functional environments that enable users to find comfort and accessibility in the information provided in order to favor this mode of learning.

## Future lines of work

The potential of technologies adapted for multimodal interaction in language testing offers huge possibilities for development of innovative applications. In that sense, devices will enable users to select between using one mode or another exclusively (for example, using an online dictionary or making a voice call), to the possibilities of changing between modes of interaction in the same session (sequential multimodality, as in consulting the dictionary on occasions on a mobile during a test), to true freedom in combining and changing modes (simultaneous multimodality: talking, keying, dialing, viewing, etc.) on terminals or ubiquitous devices that enable simultaneous access to voice and data channels, and thus offers opportunities for new items that resemble more what speakers do with the language and how they use it.

In reference to the project impact in the Spanish educational system, it is believed that an inexpensive system to assess speaking skills may have two potential benefits: first, it will enable testing of this skill at a low cost; and second, the impact on the classroom of implementing speaking skills may lead to a great educational improvement in foreign languages. Thus, as a whole, the expected effect of the project if used in the near future is immense and certainly very significant for the educational system.

To conclude, while the use of mobile phones for high stakes testing may be feasible, it is necessary to obtain a commitment from all the stakeholders including the students and the administration authorities. Since the oral test is a social, professional, and educational demand, delivering the oral section of the PAU through mobile phones would require adequate facilities from all the high schools, a better understanding of technology from teachers and new ways to plan and prepare for the test on the part of students. Researchers should also seek ways to overcome the difficulties associated with hearing impairment or other restrictions. While mobiles could be a great asset in education, it is necessary to recognize that not all teachers may be equally prepared to face such a technological change or eager to change their ways of teaching to cater to the students' needs by facilitating them with the necessary strategies for taking the test. Thus, practitioners should also receive the necessary instructions and courses to facilitate their adaptation to the new context. Nevertheless, it is believed that this change would not be any more traumatic than others that they have seen in recent years. The ongoing work in the PAULEX project is expected to continue to address these issues. The information obtained so far, while initial, provides enough evidence for the potential of this innovation in both the national and international contexts in areas such as educational planning, course design, test delivery, specifications, and information and communication technologies development. It also takes the use of mobile phones far beyond their traditional perspective of mere supportive elements of courses or learning to enhance their role as high stakes testing facilitators.

## Acknowledgements

The authors would like to thank the Ministry of Economy and Competitiveness for funding the research project (with co-financing by ERDF) within the framework of the National R + D + I (2011-2014) "Guidance, proposals and teaching for English section in the entrance examination to the University" (Reference FFI2011-22442). The researchers would also thank the Spanish Ministry of Education, Culture, and Sports because without the grant for the Senior Researchers Mobility this paper would probably have not been possible.

## References

- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.
- Baldwin, D., Fowles, M., & Livingston, S. (2008). *Guidelines for constructed responses and other performance assessments* (No. RR-07-02/ TOEFL iBT 02). Princeton, NJ: Educational Testing Service.
- Canale, M. (1984). A communicative approach to language proficiency assessment in a minority setting. In C. Rivera (Ed.), *Communicative competence approaches to language proficiency assessment: Research and application* (pp. 107-122). Clevedon, UK: Multilingual Matters.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*, 1-47.

- Chae, M., & Kim, J. (2004). Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour & information technology*, 23(3), 165–181.
- Custer, R. L., Scarella, J. A., & Stewart, B. R. (1999). The modified Delphi technique—A rotational modification. *Journal of Vocational and Technical Education*, 15(2), 50–58.
- Downey, R., Farhady, H., Present-Thomas, R., Suzuki, M., & Van Moere, A. (2008). Evaluation of the usefulness of the Versant for English test: A response. *Language Assessment Quarterly*, 5, 160–167.
- García Laborda, J. (2012). Preliminary findings of the PAULEX project: A proposal for the Internet-based valencian university entrance examination. *Journal of language teaching & Research*, 3(2), 250–255.
- García Laborda, J., & Giménez López, J. L. (2010). Aplicaciones ubicuas para la evaluación de lenguas extranjeras a través de dispositivos móviles: Evolución del proyecto PAULEX. [Ubiquitous applications for the assessment of foreign languages through mobile devices: Evolution of the PAULEX project]. In M. R. Caballero Rodríguez & M. J. Pinar Sanz (Eds.), *Modos y formas de la comunicación humana* (Vol. 2, pp. 941–946). Ciudad Real, Spain: Universidad de Castilla La Mancha.
- García Laborda, J., Magal-Royo, T., de Siqueira Rocha, J. M., & Álvarez, M. F. (2010). Ergonomics factors in English as a foreign language testing: The case of PLEVALEX. *Computers and Education*, 54(2), 384–391.
- García Laborda, J., Giménez López, J. L., & Magal Royo, Teresa. (2011). Validating mobile devices in the Spanish University Entrance Exam English paper. *The new Educational Review*, 25(3), 160–170.
- Gimenez López, J. L., Magal-Royo, T., Garde Calvo, F., & Prefasi Gomar, S. (2009). The adaptation of contents for the creation of foreign language learning exams for mobile devices. *International Journal of Interactive Mobile Technologies (iJIM)* 3, 15-17. Retrieved April 30, 2014 from <http://online-journals.org/i-jim/article/view/1008>
- Giménez López, J. L., Magal Royo, T., García Laborda, J., & Garde Calvo, F. (2009). Methods of adapting digital content for the learning process via mobile devices. *Procedia - Social and Behavioral Sciences*, 1(1), 2673–2677.
- Keskin, N. O., & Metcalf, D. (2011). The current perspectives, theories and practices of mobile learning. *TOJET: The Turkish Online Journal of Educational Technology*, 10(2), 202–208.
- Magal-Royo, T., & Gimenez-López, J. L. (2012). Multimodal interactivity in the foreign language section of the Spanish university admission examination. *Revista de Educacion*, 357, 163–176.
- Magal-Royo, T., Fajarnes, P., Tortajada Montañana, G. I., & Defez Garcia, B. (2007). Evaluation methods on usability of M-learning environments. *International Journal of Interactive Mobile Technologies (iJIM)*, 1(1). Retrieved August 10, 2013 from <http://online-journals.org/i-jim/article/view/148>
- Magal-Royo, T., Giménez-López, J. L., Paíry, B., García-Laborda, J., & Gonzalez-Del Rio, J. (2011). Multimodal application for foreign language teaching. *Proceedings of the 14th International Conference on Interactive Collaborative Learning: ICL 2011 - 11th International Conference Virtual University* (pp. 145–148). Kassel, Germany: Kassel University Press.
- Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3–67.
- Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6–20.
- Nelly, B., Nevile, L., Sloan, D., Fanou, S., Ellison, R., & Herrad, L. (2009). From Web accessibility to Web adaptability. *Disability and Rehabilitation: Assistive Technology*, 4(4), 212–226.
- Nielsen, J., & Mack, R. L. (1994). *Usability inspection methods*. New York, NY: John Wiley & Sons.
- Oviatt, S., & Cohen, P. (2000). Multimodal interfaces that process what comes naturally. *Communications of the ACM*, 43(3), 45–53.
- Oviatt S., & Larson, J. A. (2003, April). *Principles for multimodal user interface design*. Paper presented at the Workshop of the Conference on Human Factors in Computing Systems. Extended Abstracts on Human Factors in Computing Systems CHI EA '03. ACM New York, NY, USA 1058-1059 doi: 10.1145/765891.766147 (Consulted 29/04/2014).
- Oviatt, S. (1999). Ten myths of multimodal interaction. *Communications of the ACM*, 42(11), 74–81.
- Park, Y. (2011). A pedagogical framework for mobile learning: Categorizing educational applications of mobile technologies into four types. *International Review of Research in Open & Distance Learning*, 12(2), 78–102.

- Piolat, A., Roussey, J. Y., & Thunin, O. (1997). Effects of screen presentation on text reading and revising. *International Journal of Human-Computer Studies*, 47(4), 565–589.
- Poehner, M. E. (2008). *Dynamic assessment: A Vygotskian approach to understanding and promoting second language development*. Berlin, Germany: Springer Publishing.
- Salmani-Nodoushan, M. (2009). Measurement theory in language testing: Past traditions and current trends. *Journal on Educational Psychology*, 3(2), 1–12.
- Sariola, J. (2003). The boundaries of university teaching: Mobile learning as a strategic choice for the virtual university. In H. Kynaslahti & P. Seppala (Eds.), *Mobile Learning* (pp. 71–78). Helsinki, Finland: IT Press.
- Valk, J. H., Rashid, A. T., & Elder, L. (2010). Using mobile phones to improve educational outcomes: An analysis of evidence from Asia. *The International Review of Research in Open and Distance Learning*, 11(1), 117–140.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, Massachusetts: Harvard University Press.
- Weining, Y., Heng, W., & Guoping, W. (2007). Designing transparent interaction for ubiquitous computing: Theory and application. In J. A. Jacko (Ed.), *Human-computer interaction. Interaction design and usability* (pp.331–339). Berlin, Germany: Springer.
- Weir, C. J. (2005). *Language testing and validation*. Basingstoke, UK: Palgrave Macmillan.