Journal of Universal Computer Science, vol. 22, no. 1 (2016), 114-123
submitted: 30/6/15, accepted: 20/12/15, appeared: 1/1/16 © J.UCS

Looking towards the Future of Language Assessment: Usability of Tablet PCs in Language Testing

Jesús García Laborda
(Universidad de Alcalá, Alcalá de Henares, Spain
jesus.garcialaborda@uah.es)

Teresa Magal Royo
(Universitat Politecnica de Valencia, Valencia, Spain
tmagal@degi.upv.es)

Margarita Bakieva
(Universitat de Valencia, Valencia, Spain
margaritabakieva@gmail.com)

Abstract: This research addresses the change in how the Spanish University Entrance Examination can be delivered in the future. There is a wide acknowledgement that computer tests are very demanding for the delivering institutions which makes computer language testing difficult to implement. However, the use of tablet PCs can facilitate the delivery at even lower cost than the regular computer based language testing. 183 students in their last year **of** high school took a computer based language test which included reading, writing, listening and speaking. The research aspects were 1) whether they feel at ease with the tablet PC exam; 2) if they felt that visual aspects were accessible and 3) whether the interface organization was clear. The paper first has a brief description of the OPENPAU® platform, after it addresses the field study based on questionnaires and observations with students during the test delivery, finally the results indicate that this means of language test delivery could be adequate not only for the Spanish University Entrance Examination but for most standardized tests.

Keywords: mobile assisted language learning, higher education, standardized testing, language learning, college entrance examination
Categories: J.4, J.5, L.3.6

1 Introduction

E-learning has become a prevailing move in the last 20 years. There are a number of reasons to explain its increasing importance among which are low-cost delivery, immediacy of results, increasing familiarity with computers and mobile devices for education, and others. E-testing, however, is not increasing accordingly. There are many reasons but among them one of the most significant aspects is the necessary control over test context and location (the room where the test is taken), the necessary standardization of testing platforms [Roever, 01] [García Laborda, 02] [Garcia Laborda, 03] to ensure validity and security. Additionally, there is a significant difficulty in creating the adequate digital contents to serve as prompts [Gimenez Lope et al., 04]. As a consequence, despite the increasing demand for this type of tests, the implementation has been rather slow. The most significant language tests are

administered by large organizations such as the Educational Testing Service (ETS) which administers the TOEFL (Test of English as a Foreign Language), TOEIC (Test of English for International Communication), Graduate Record Examination (GRE) General and Subject Tests, or the British Council with the Cambridge Suite and the International English Language Testing System (IELTS). However, up to this point, these tests are delivered in specially designed locations with very specific setting technical conditions. This is thus a major constraint for state exams worldwide which are usually delivered at a much lower expense and usually supervised by regular teachers and not by professional examiners. Therefore, one significant need is providing educational administrators with a valuable tool that can be accessible and adaptable to more general conditions than those met by testing centers run by ETS, Pearson® or the British Council®.

Ubiquitous learning is currently an adequate mode of facilitating test delivery in a range of different contexts both across the country and within an educational setting. For instance, mobile devices permit rural schools to be reached even in the most distant locations and also make available a large number of rooms / classes while for traditional desktops only a limited number of computer rooms or computer labs can be used for e-learning [Kukulska-Hulme, 05] or to deliver tests. This has positive effects on standardized tests because more schools can serve for testing and also because, in not a few cases, students do not need to move to external testing centers (in Spain this distance can be up to 100 km.).

Familiarity with mobile devices is also a significant asset in this kind of delivery. These applications are of common use for both ludic and learning aspects. While it is debatable if most students use it for educational purposes, certainly they are familiarized on how to operate them and especially how to get the best from them. They use these devices for communication and most relevant aspects of their lives. Thus, the old paradigm of technology familiarity has been overcome for the younger potential testees.

The main goal of this innovative research was to investigate whether participating students can adapt to the use of mobile devices for language testing. As a consequence, the general aim of this research is to foresee if technical issues can arise from the use of tablet PCs in language testing. Therefore, this research intends to contribute to the comprehension of the key matters in the future implementation of similar tests not only in Spain but worldwide. In this sense, this study with its limitations is an in-depth first approach to a matter which may be familiar in a near future.

The research aspects were 1) whether they feel at ease with the tablet PC exam; 2) if they felt that visual aspects were accessible and 3) whether the interface organization was clear.

2 User adaptation through multimodal navigation

As mentioned in the introduction, in language testing students need to feel at ease or, at least, not feel that the effect of recording their speaking will have a negative effect on their performance. The research team considered that the behavior of a computer while taking a test should be agile, easy and verifiable (and reliable in recording the testee's performance). Obviously, although the students already know the objectives and how the exam will be like on pen-and-paper, things can be extremely different when they have to face an interface full of icons, supported by images and texts where notions of navigability are necessary. In fact, even today computers cannot be said to have an invisible effect in language testing while their move in e-learning is more and more evident each day. Thus, the interface will necessarily be aimed not only to the final users but also to the task to be deployed.

Mobile devices usually meet a set of functional and aesthetic criteria that have a direct influence on the user's response to a certain task. In the case of a foreign language exam where tasks include speaking and writing in a foreign language, mobile devices require an interface adaptation so that the visual elements can be recognized easily [Magal-Royo et Al., 06]. That is the reason why research on the adaptation of graphic textual contents in mobile devices progressively focuses not only on a usable interface but on the development of a multimodal interaction that enables the choice (whether through the keyboard or the microphone) that best adapts to the testee's needs and facilitates navigation.

Given the difficulties to achieve the universal accessibility for the user, currently the idea of free choice of mode of interaction device-user has been developed. For instance, if a testee has some kind of visual problem he may want to concentrate on the test itself rather than on navigation and would prefer using his voice for such operation [Sharma et Al. 07]. In this way, multimodality in computer based language testing obviously also enhances the power of user-device synchronous interaction [Oviatt and Cohen, 08] [Oviatt and Larson, 09].

Until lately, the most common means of computer-user interaction where the keyboard and the mouse, and in the last few years haptic sensors, tactile screens and voice recognition among others, are helping to facilitate the user interaction with the computer and the versatility between user and the mobile device is one the cornerstones in technical accessibility for future educational devices [Magal-Royo et al., 10].

In the OPENPAU® project, the research team considered that students should have, at least three multimodal ways of interaction for tablet PCs: voice recognition (currently in revision after the first trials), touch screen, and keyboard for the large composition. Therefore, the research team considered that this evolution towards an architecture based on a multimodal organization was an asset which has never been implemented [Magal-Royo and Giménez Lopez, 11]. The OPENPAU® platform for language testing incorporates in its user-aimed architecture three main aspects:

1) The possibility of selection from three main means of interaction for navigation at the time of the test (keyboard, mouse and voice) (Figure 1).
2) Use of an interface with guided screens and online access based on an adaptable multiform navigator for different types of devices (including both mobile devices and desktop computers).

3) The use and adaptation of Moodle® (Open Source) because of its broad use worldwide by many educational institutions (like the Spanish Ministry of Education).

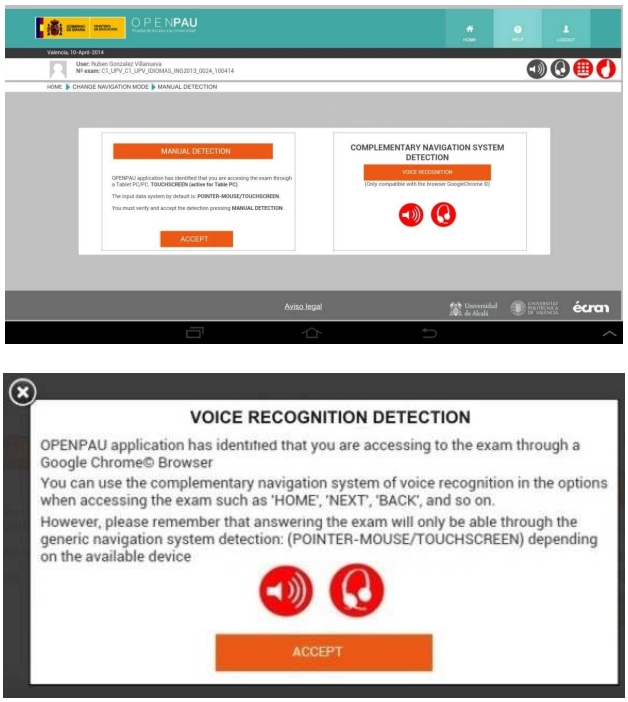

Figure 1: Selection of user-centered navigation mode in the OPENPAU® platform.

3 The OPENPAU® project of Mobile Language Testing

From Chapelle's perspective [Chapelle, 12] testing serves to obtain evidence to support inferences about the testee's language competence. In the case of the Spanish University Entrance Examination (PAU henceforth) there are no formal rubrics but [García Laborda, 03] suggests that the foreign language section should serve to know whether students will have the language skills to favor their academic professional development. In this sense, mobile devices have an importance both linguistically and as a means of information retrieval. In this sense, a mobile test would meet standards of validity of delivery and reliability.

The OPENPAU® project started in 2010 as a way to find the adequate delivery means (whether face to face to technology based) [Garcia Laborda, Magal Royo, Litzler & Gimenez Lopez, 13]. The platform (figure 2) has four sections. Each corresponds to one specific skill (reading, writing, listening and speaking). The test comprises the following parts:

a. Reading. A text followed by three main types of questions: multiple choice, True/False selection, and matching questions and responses.
b. Writing. An essay of variable length.
c. Listening. Three podcasts followed by the types of questions mentioned for listening: multiple choice, True/False selection and matching questions and responses.
d. Speaking. A video followed by four mini-clips where the students respond to three short questions and a longer two-minute longer speech (also prepared with pen and paper).

The OPENPAU® platform is especially attractive because it permits the user to select the type of interaction (keyboard, mouse or voice recognition) for the online test on a Chrome® browser. Thus, testees can progress or go back in the text by using voice recognition. As a consequence, navigability has been increased greatly by diminishing certain potential and physical handicaps that some students can have. However, voice recognition in this platform is still to be further developed.

The programming language used is based on source code specifically developed for Moodle©. The use of Moodle was considered for its flexibility because it permits generation and adaptation of contents not only for virtual classrooms (or labs) but also development of specific courses (which usually require standardized tests) and includes new tasks based on the use of sound, graphics and videos. The OPENPAU® assembly has been developed in XHTML+CSS so that it can be properly displayed on mobile devices such as tablet PCs. The OPENPAU® platform has two main differentiated environments based on Moodle©. The first is used for administration and management of users, exams, teachers and so on (backend) and the second is oriented towards the user (frontend) with a restrictive interface and formally adapted to the specifications of the PAU exam (Figure 2).

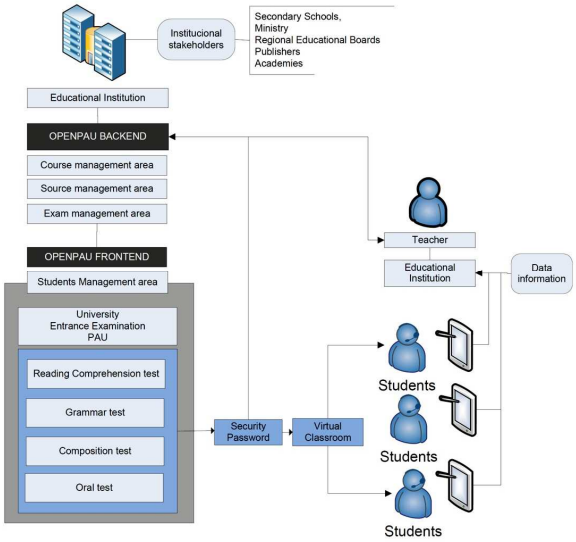

Figure 2: OPENPAU® Platform architecture

The actual field research on functionality and usability was done with a tablet PC Wolder miTab EVOLUTION W2 10.1" HD IPS reinforced, with QUAD CORE, 16 GB, and a QWERTY BT keyboard. The user's frontoffice interface has been specifically designed for a language test according to Fulcher's principles of interface design for language testing [Fulcher, 14] (Figure 3). Navigation and other visual resources (Icons, help, etc.) have been limited to a minimum to avoid technology bias. Only the center of the interface changes according to the task. Task management, length and difficulty are carried out by the administrators in the backend section.

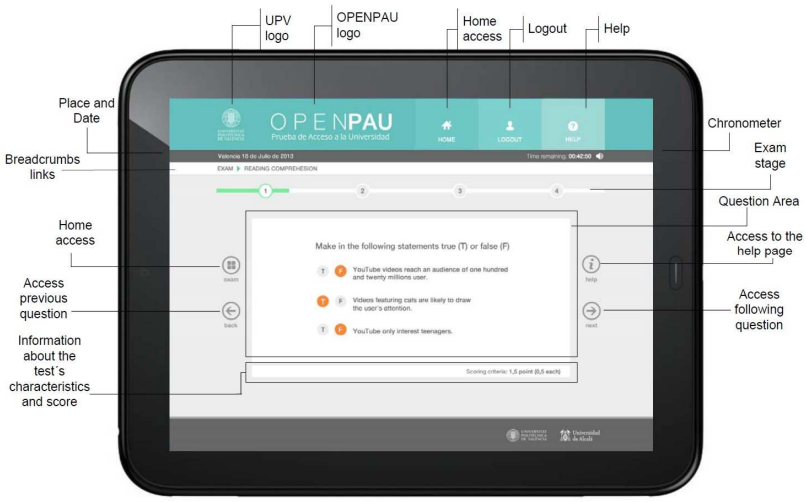

Figure 3: OPENPAU® user's interface for language testing

4 Methodology

4.1 Participants and Design

This research analyzes the participants' attitudes towards usability of the OPENPAU® platform in a test-like situation. Thus the study addresses the students' feelings as well as difficulties and advantages. To do so, the students first took the test on a tablet PC and then had the opportunity to see the version for desktop which, in fact, has a few minor differences in terms of interface but not in navigability.

The participants were 183 high-schoolers in their last year in school in Spain of different majors including Sciences, Social Sciences and Humanities from 7 high schools in Madrid and Guadalajara (Spain). The schools were located and classified as rural and urban as well as of higher and lower middle class. The research team considered that the limited number of extremely wealthy and depressed schools were not representative in these two provinces of Spain. The students had a homogenous school background and the variables of sex or social class, although important, were not considered significant at this point.

Most of these students are from Spanish Peninsular origin but the sample also included Moroccan (4), Romanian (12), Bulgarian (2), English (1), Russian (1) and Chinese (4) which is an indicator of a variety of origins but the limited influence they could have in this research. Students had taken, at least, six years in English but most 12 because this subject is compulsory in Spain between 1st and 12th grade (all the years in primary and secondary school). Their English proficiency level was from almost beginners to higher intermediate, or A2 to B2 according to *The Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. The research team did not know the students personally so they have little additional information on them.

Although most students acknowledged that they had some kind of mobile device from a laptop to a tablet PC to a smart mobile phone, 84% recognized that their use of those devices for learning was limited to browsing information and, sometimes, doing some online exercises. The use for the English class was even more limited 34%.

4.2 Data and Analytic Process

In order to investigate this issue, the research team organized experimental sessions which consisted in first taking a real past test of the English section of PAU enhanced with listening sections and videos. First they did the full version on a tablet PC and afterwards a reduced version with a desktop computer. At the end of these two, they filled in a mixed quantitative/qualitative questionnaire with their experiences.

Questions in this questionnaire were organized into three main sections: 1) Overall sense of user's satisfaction; 2) attitudes towards visual organization; and 3) attitudes towards visual elements and icons. The first and the last questions in the questionnaire asked for more general opinions of satisfaction while the other four questions intended to see if the specifics of interface and architecture design form the final user's satisfaction. The responses were ranked in a Likert-scale from one to five with one meaning "I do not agree at all" and 5 meaning "I totally agree".

5 Results

The results herby presented (table 1) are based on a selection of items from a larger questionnaire mostly revolving around usability and final impression after the use of tablet PCs. In this section we present three kinds questions: general impressions, visual organization, internal organization and further applicability.

5.1 Overall satisfaction

The research team considered that ease of use would be desirable to achieve the "invisible effect" of technology on proficiency measurement. In that sense, there is a significant difference between those who believe that this platform is easy to use and those who do not (X^2=13.738; p-value: 0.0002).

The results of this question are similar to the last questions where most students (51%) observe that the things in the platform work as they expected them to (X^2=36.83; p-value: 0).

	1	2	3	4	5
	Percentage (%)				
1.This type of interface is easy to use me	16,5	7,8	26,1	27,5	22,0
2.The visual organization seems appropriate for a student of 10 grade (4ºESO)	13,3	14,2	36,7	24,3	11,5
3.The visual organization seems appropriate for a student of 12th grade (2º Bachillerato).	12,8	10,6	23,4	25,7	27,5
4. The interface elements allow me to easily move around the application	13,3	8,7	37,2	28,0	12,8
5.The visual elements let me recognize the important parts of the interface	11,0	10,6	35,8	31,2	11,5
6.There are things that do not work as I think they should	11,0	13,3	24,8	21,6	29,4

Table 1: Results of the survey on the students' attitudes towards usability of the OPENPAU® testing platform.

5.2 Visual organization

The results in this section are even clearer although there is a slight difference. The percentage of students who believe that the interface is clear is a bit higher for the 12th graders. This may lead to considering that students in 12th grade are probably more used to working with computers than their counterparts or they think that they have just evolved in the use of computers as learners. This can be especially seen in the "5" value which is 14% higher for 12th graders than for 10th graders. It is also worthy of note to see the degree of indecision in their consideration about tenth graders, especially because that experience is not so distant in their lives.

5.3 Visual elements

A similar result can be seen in relation to the layout of the visual elements. Most students agree that the elements enable easy visual orientation (40% agree versus 23.4%) and that they show clearly what the important parts in the platform are (42.7% versus 21.6%). These two questions also show that a little more than one third of the students did not have a clear answer and, hence, did not position themselves on one side or the other.

6 Conclusions

This research intended to investigate the students' attitudes and acceptance of a new platform for language testing for the PAU. This study displays a homogeneous picture of computer based language testing through tablet PCs. Although these students had never had any direct experience with computer assisted language testing, they still showed evidence of ability and did not show any rejection to being tested. However, the influence of the invisible effect of the delivery means is still to be debated for the further research. The responses clearly indicate that the suggested design could be valid but, as one of the limitations of this research, the reduced set of questions in relation to usability necessarily carries the need to continue researching in this sense, especially if the Ministry of Education or the regional boards are to finally implement the test on computers or, even better, on tablet PCs.

The findings of this research also facilitate the understanding of the students' preparation towards the implementation of the test. According to these results, students would be able to take the test without any sense of distress due to the use of technology. Thus, students' attitude towards technology integration, one of the main issues in computer language testing, would probably not have a significant initial effect. As mentioned, however, this notion should also be supported by more empirical evidence.

7 Future Work

Further research includes the comparability of language test results with similar tests in pen and paper format. Nevertheless, if technology is initially not a constraint, despite the similarity of computer and pen-and-paper tasks, the increased richness of the new tasks could have a potential effect which has generally been considered positive but that needs specific study in the students' context and realities. Thus, more evidence is required before claiming the success of this means of test delivery. Still, even if the European Union is moving towards computer testing, educational administrators must be willing to make an initial financial effort in terms of software, hardware or staff training. Thus, additional research needs to be done to seek the real cost and benefits of implementing this new format for the PAU exam.

Acknowledgements

The researchers would like to express their gratitude to the Ministry of Research and Innovation of Spain (MICINN) for supporting the development and implementation OPENPAU research project (FFI2011-22442) with cofounding with FEDER funds under the 2008-2011 plan.

References

[Chapelle, 12] Chapelle, C.: "Validity argument for language assessment: The framework is simple", Language Testing, 29 (1) (2011), 19-27.

[Council of Europe, 01] Council of Europe: The Common European Framework for Languages: Learning, Teaching, Assessment (2001)
http://www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf

[Fulcher, 14] Fulcher, G.: "Interface design in computer-based language testing", Language Testing, 20 (4) (2003), 384 - 408.

[García Laborda,2] García Laborda, J.: On the Net: Introducing Standardized EFL/ESL Exams, Language Learning & Technology, 11 (2) (2007), 3 - 9.

[García Laborda, 03] García Laborda, J.: "¿Necesitan las universidades españolas una prueba de acceso informatizada? El caso de la definición del constructo y la previsión del efecto en la enseñanza para idiomas extranjeros", Revista de orientación y Psicopadagogía, 21(1) (2010),71-80.

[García Laborda et al.,13] García Laborda, J., Magal Royo, T., Giménez López, JL.: Frances Litzler M.: "Mobile phones for a University Entrance Examination language test in Spain", Educational Technology & Society 17 (2) (2014),17-30.

[Gimenez Lopez et al., 4], Giménez López, J.L., Magal Royo, T. García Laborda, J., Garde Calvo, F.: "Methods of adapting digital content for the learning process via mobile devices", Procedia - Social and Behavioral Sciences, 1, (2010), 2673-2677.

[Kukulska-Hulme, 5] Kukulska-Hulme, A.: "Re-skilling language learners for a mobile world", The International Research Foundation for English Language Education. (2013), Monterey, CA http://www.tirfonline.org/english-in-the workforce/mobile-assisted-language-learning/

[Magal-Royo et Al., 6] Magal-Royo T., García Laborda J., Peris-Fajarnes G. and Spachtholz, P.: "Visual Learning Through Guided Iconography in Wireless Scenarios", Proceedings of ECEL 2007. 6th European Conference on e-learning, (2007), 415-418.

[Magal-Royo et Al., 10] Magal-Royo, T., Laborda, Garcia J. Giménez-Lopez, J. L. Bakieva, M.: "Accessible Multimodal Interaction for Language Learning on Mobile Devices", International Conference on Applied Social Science, ICASS 2011 (2) (2011), 47-51.

[Magal-Royo and Giménez Lopez, 11], Magal-Royo, T. and Giménez Lopez, J. L.: "Multimodal Interactivity in the Foreign Language Section of the Spanish University Admission Examination", Revista Educación 357 (2012), 163-176.

[Oviatt and Cohen, 08] Oviatt S., Cohen P.: "Multimodal interfaces that process what comes naturally". Communications of the ACM, 43 (3) (2000), 45-53.

[Oviatt and Larson, 09], Oviatt S. and Larson J.A.. "Principles for Multimodal User Interface Design", Workshop. CHI 2003: New Horizons (2003),1058-1059.

[Roever,1] Roever, C.: "Web based language testing. Language Learning and Technology", 5 (2) (2001), 84 - 94.

[Sharma et Al. 7]. Sharma R., Pavlovi V.I. and Huang T. S.: "Toward Multimodal Human-Computer Interface", Invited paper at Proceedings of the IEEE, 86 (5) (1998), 853-869.