24 Using tablet PC's for the final test of Baccalaureate

Jesús García Laborda[1] and Teresa Magal Royo[2]

Abstract

Online testing is becoming a popular way to deliver language tests, partly because of its reduced cost, partly because of the high quality of test data collection. In language tests, interface validation has received a limited attention in professional literature (García, Magal, da Rocha, & Fernández, 2010). This paper will show the validation process of the OPENPAU application, which aims at language testing exam delivery through an attitudinal study. The real importance of this project is that it offers reliable solutions at a low cost for the needs of a nationally delivered online test that can serve to assess all the traditional language skills (namely speaking, listening, reading and writing) in an efficient, simple and cost-effective manner. Our paper presents the design, development, and evaluation of a tablet PC software application for assessing both productive and receptive skills in foreign languages for its prospective use in the Final Test of Baccalaureate. The software development has been carried out within the OPENPAU project (FFI2011-22442), funded by the Ministry of Economy and Competitiveness.

Keywords: computer system design, computer software, usability, computer assisted testing, internet, computer software evaluation, educational technology.

1. Universidad de Alcalá, Alcalá de Henares, Madrid, Spain; jesus.garcialaborda@uah.es

2. Universitat Politècnica de València, Valencia, Spain; tmagal@degi.upv.es

1.   Introduction

The English section of the University Entrance Examination (PAU) is a challenge in Spain even today. A number of projects and studies have intended to show some of its main assets and pitfalls. Fernández and Sanz (2005) probably did one of the best diachronic studies on the topic. Their paper reflected many of the different sound research articles that it included. Whittaker (2006) mentioned that there are some historical reasons why the foreign language paper was included in the PAU. It is also clear that the revision of the Foreign Language section of the University Entrance Examination seems totally necessary.

In 2009, the Ministry of Education suggested the introduction of a number of tasks by 2011. In that year, the Ministry of Education, Culture and Sports decided to postpone those changes to 2013 and later the LOMCE was passed and finally the exam will disappear in 2017 (Ministry of Education, Culture & Sports, 2013). However, despite the experts' opinions, as years go by, it is also self-evident that the different national and regional governments have no interest to improve an old fashioned test that can hardly provide any information on the student's competence (Amengual, 2005, 2006; García, 2010, 2012). It does not matter whether newer ideas have been provided to increase the validity of the test. In this sense, a few years ago Catalunya began to deliver the listening comprehension tasks, later Galicia did the same. Very few studies, however, have addressed whether this change has actually had any impact in the classroom (Wall, 2000). What certainly seems true is that the level required to Catalan students is higher than that demanded of the rest of the Spanish students. Thus, maybe the inclusion of certain tasks has the power to change the educational approach in language testing (Alderson & Wall, 1996; Wall, 2005).
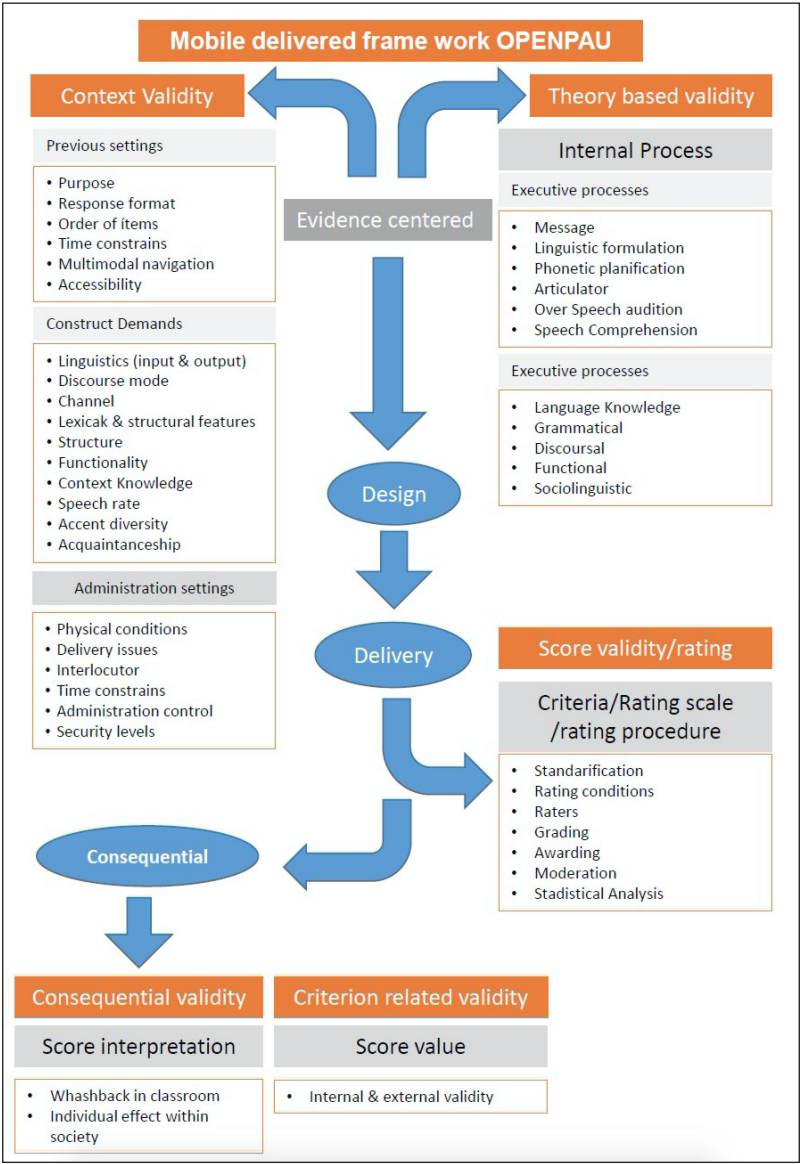
2.   Literature review

The effect of tests has been long documented in professional literature (Knudson, 1975; Kohonen & Nummenmaa, 1976; Messick, 1996; Pimsleur,

1975; Shohamy, 1992 – among many others). The concern in these papers was on the negative effects on teaching by jeopardising language learning to give ground for test scores and accountability. In 1989, Hughes defined backwash as "[t]he effect of testing on teaching and learning" (p. 1). Spolsky (1994) considers that backwash relates to the side-effects of testing in the classroom. One of the first documented changes introduced by positive washback is described by Wall (2005), who performed a total change in education in Sri Lanka in accordance to a totally new high school leaving exam. This experience helped to create a positive construct by Alderson and Wall (1993). According to Weir (2005), this construct does not just include the task rubrics, but a much larger set of variables and conditions in which the way of delivery has a significant role (Figure 1).

According to Weir (2005), all the factors included in the context validity group may have a potential effect in changing the results of a test. The theory also states that the use of different means of delivery, however, should have a minimal effect in the final score. Indeed, the documented number of learning experiences on the use of tablet PC limits the capacity to discern whether they have a potential positive or negative affect (Chapelle, 2001). According to van Oostveen, Muirhead, and Goodman (2011), there is little evidence to support that comfort and familiarity with a tablet PC or iPad. For them, "[i]t is important to note that the introduction of a new technology, even if it makes a wide variety of affordances available for use, cannot by itself instigate redefinition of learning tasks to allow for meaningful learning to occur" (van Oostveen et al., 2011, p. 78). However, Crichton, Pegler, and White (2012) consider that the use of a tablet PC requires the familiarisation with the purpose of use as well as with the new educational technology. This also means that although students might be familiarised with the use of mobile technology for leisure and social networking, its use for educational purposes (especially for testing) may seem unusual or even odd for them. Similar results are found by Waters (2010), who considers that iPads may be way behind in usability versus netbooks or even laptops, especially in writing (Sullivan, 2013). Although some of these studies remark the idea of students' limitations to use traditional tablets, up to now, no studies have addressed the use of tablets with an external keyboard.

Figure 1. Theoretical framework of the OPENPAU project (based on García, Magal, Litzler, & Giménez, 2014 and Weir, 2005)
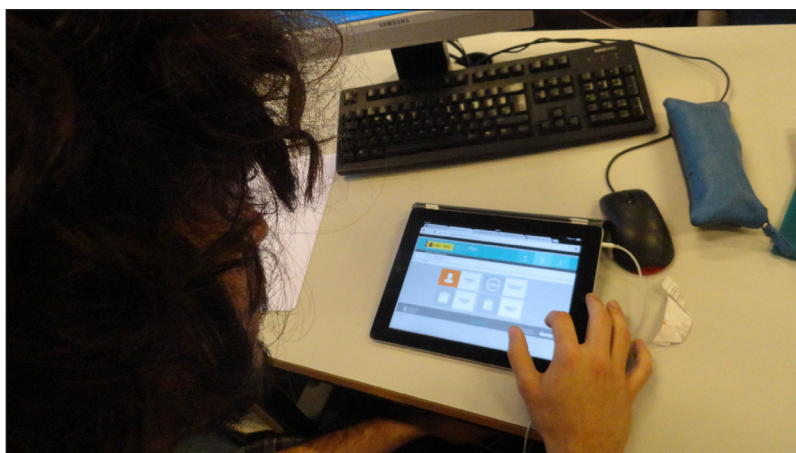
As for their use in testing, there are very few. For instance, Siozos, Palaigeorgiou, Triantafyllakos, & Despotakis (2009) said that the use of tablets in their experiment in secondary education proved their validity. On the contrary, in one of the most interesting papers on the topic, Schaffhauser (2012) states that it is necessary to consider "the challenges of security, usability, and content that might arise when students are taking tests on tablet devices and discussing how these might be resolved" before sound high stakes exams can be delivered by tablet or iPad (p. 16).

3.   Methodology

In order to observe the degree of satisfaction of using a tablet PC as a delivery system for the University Entrance Examination, an online questionnaire was delivered to 31 first year university students at Universidad Politécnica de Valencia in January 2014, with a primary version of the OPENPAU prototype of tablet-delivered online test. The test is intended to be taken with 10 inch tablets with a Windows 8 environment (see Figure 2).

Figure 2.   Student taking the tablet-based language test through the OPENPAU application
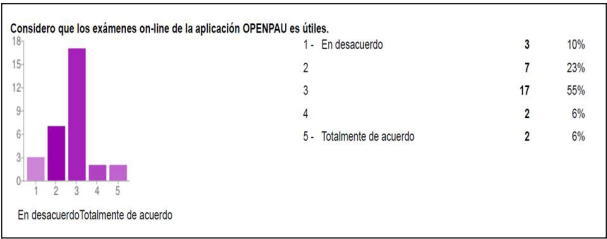
4.   Results
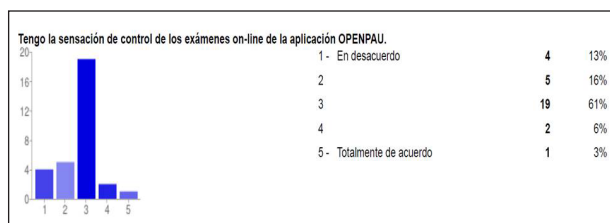
The questionnaire revolved around three main items.

1) Considero que los exámenes on-line de la aplicación OPENPAU son útiles (I consider that the online tests delivered through the OPENPAU application are useful). The results indicated that students were mostly sceptical towards the use of this type of test (Figure 3). This response can be due to the lack of previous experience. Low-stakes tests are not usually delivered online, and the only high-stakes that most students may have had to take in their school life in Spain is the PAU, which obviously is not computer-based. As a consequence, many may feel strange when taking an online foreign language test. Another issue could be their own attitude towards how foreign languages should be assessed. A third reason for this response would be the lack of specific preparation and skills to take such tests despite its user-friendly design and usability.

Figure 3.   Response to 'I consider that the online tests delivered through the OPENPAU application are useful'
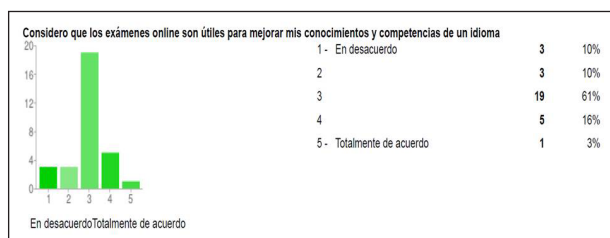

2) Tengo la sensación de control de los exámenes online de la aplicación OPENPAU (I have the sensation of control of the online tests of the OPENPAU application). The responses to this item indicate the same tendency as in the previous item (Figure 4). The most evident indication is that there is a great number of undetermined responses and a slight tendency towards a negative feeling of control. Test taker training may be the key issue in this case. It is important to note that participants in this research had never had any contact with the application before. Thus, their negative responses could have been expected.

Figure 4.   Response to 'I have the sensation of control of the online tests of the OPENPAU application'

Tengo la sensación de control de los exámenes on-line de la aplicación OPENPAU.

1 - En desacuerdo	4	13%
2	5	16%
3	19	61%
4	2	6%
5 - Totalmente de acuerdo	1	3%

3) Considero que los exámenes online son útiles para mejorar mis conocimientos y competencias de un idioma (I consider that online tests are useful to improve my knowledge and competence in the language). Item 3 may contrast with Item 1, since students were more positive towards using tests than to the platform. However, although the results are slightly better, no significant conclusions could be drawn. As in the two items before, the largest response was the subjects' inability to take a position and the central tendency was the largest response (Figure 5).

Figure 5.   Response to 'I consider that online tests are useful to improve my knowledge and competence in the language'

Considero que los exámenes online son útiles para mejorar mis conocimientos y competencias de un idioma

1 - En desacuerdo	3	10%
2	3	10%
3	19	61%
4	5	16%
5 - Totalmente de acuerdo	1	3%

En desacuerdo   Totalmente de acuerdo

5.   Conclusions

Although the results of this short study did not evidence the interest of the students in this application, given the current tendency of research in the use

of tablets for education, we consider it of high interest at a national level to continue with the current research. This work is just a first approach towards this issue but, as indicated by Item 3, if the adequate conditions are present, students could be receptive to tablet-based language tests and recognise the value of online-delivered language tests. In this sense, given the current state of affairs in contemporary education, it would be interesting to study responses to similar studies in a few years. Besides, further studies should focus on the other stakeholders' (teachers, administrators, educational boards and so) attitudes towards the implementation of this delivery system. Additional studies should study the impact of training on the new tasks, the presence of underdeveloped skills (especially speaking) in instruction and organisation, and administration issues.

This paper provided the theoretical framework and a very brief description of the OPENPAU tool. Further publications will provide more detailed information on the importance for ubiquitous language learning through training, advantages of mobile based-testing, and types of semi-assisted speaking interaction with tablet-based delivery. The use of tablets in the Final Test of Baccalaureate will depend to a large extent on budgets and the whole framework of this exam in the future. At this point, our hopes are high, but only time will tell.

6.   Acknowledgements

References

Alderson, C., & Wall, D. (1993). Does washback exist? *Applied Linguistics, 14*(2), 115-129. Retrieved from http://dx.doi.org/10.1093/applin/14.2.115

Alderson, J. C., & Wall, D. (1996). Examining washback: the Sri Lankan impact study. In A. Cumming & R. Berwick (Eds.), *Validation in language testing*. Clevedon, England: Multilingual Matters.

Amengual, M. (2005). Posibles sesgos en el examen de Selectividad. In H. Herrera Soler & J. García Laborda (Eds.), *Estudios y Criterios para una evaluación de calidad* (pp. 121-148). Valencia: Universidad Politécnica de Valencia.

Amengual, M. (2006). Análisis de la prueba de inglés de Selectividad de la Universitat de les Illes Balears. *Ibérica, 11*, 29-59.

Chapelle, C. A. (2001). *Computer applications in second language acquisition*. Cambridge: Cambridge University Press. Retrieved from http://dx.doi.org/10.1017/CBO9781139524681

Crichton, S., Pegler, K., & White, D. (2012). Personal devices in public settings: lessons learned from an iPod Touch/iPad project. *Electronic Journal of e-Learning, 10*(1), 23-31.

Fernández, M., & Sanz, I. (2005). Breve historia del examen de Selectividad. In H. Herrera Soler & J. García Laborda (Eds.), *Estudios y criterios para una Selectividad de calidad en el examen de Inglés* (pp. 19-26). Valencia: Editorial Universidad Politécnica de Valencia.

García, J. (2010). ¿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),71-80.

García, J. (2012). De la Selectividad a la Prueba de Acceso a la Universidad: Pasado, presente y un futuro no muy lejano. *Revista de Educación, 357*, 17-28.

García, J., Magal, T., da Rocha, J. M., & Fernández, M. (2010). Ergonomics factors in English as a foreign language testing: The case of PLEVALEX. *Computers & Education, 54*(2), 384-391. Retrieved from http://dx.doi.org/10.1016/j.compedu.2009.08.021

García, J., Magal, T., Litzler, M. F., & Giménez, J. L. (2014). Mobile phones for a University Entrance Examination language test in Spain. *Educational Technology & Society, 17*(2), 17-30.

Hughes, A. (1989). *Testing for language teachers*. Cambridge and New York: Cambridge University Press.

Knudson, R. L. (1975). Emphasis: use and misuse of standardized testing. *The English Record*, *26(2)*, 114.

Kohonen, V., & Nummenmaa, L. (Eds.). (1976). Special issue on teaching and testing communicative competence. *Language centre news, 4*.

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256. Retrieved from http://dx.doi.org/10.1177/026553229601300302

Ministry of Education, Culture & Sports. (2013). *Ley Orgánica 8/2013, de 9 de diciembre, para la mejora de la calidad educative* (p. 97858-97921). Madrid, Spain: Boletín Oficial del Estado. Retrieved from https://www.boe.es/boe/dias/2013/12/10/pdfs/BOE-A-2013-12886.pdf

Pimsleur, P. (1975). Criterion vs. norm-referenced testing. *Language Association Bulletin*, *27*(1), 21-24.

Schaffhauser, D. (2012). You may now open your test tablets. *T.H.E.Journal*, *39*(7), 16-23.

Shohamy, E. (1992). Beyond performance testing: a diagnostic feedback testing model for assessing foreign language learning. *The Modern Language Journal, 76*(4), 513-521. Retrieved from http://dx.doi.org/10.1111/j.1540-4781.1992.tb05402.x

Siozos, P., Palaigeorgiou, G., Triantafyllakos, G., & Despotakis, T. (2009). Computer based testing using "digital ink": Participatory design of a tablet PC based assessment application for secondary education. *Computers & Education, 52*(4), 811-819. Retrieved from http://dx.doi.org/10.1016/j.compedu.2008.12.006

Spolsky, B. (1994). Conditions for second language learning in Israel. *English Teacher's Journal*, *47*(May), 45-54.

Sullivan, R. M. (2013). The tablet inscribed: inclusive writing instruction with the iPad. *College Teaching, 61*(1), 1-2. Retrieved from http://dx.doi.org/10.1080/87567555.2012.700339

Van Oostveen, R., Muirhead, W., & Goodman, W. M. (2011). Tablet PCs and reconceptualizing learning with technology: a case study in higher education. *Interactive Technology and Smart Education, 8*(2), 78-93. Retrieved from http://dx.doi.org/10.1108/17415651111141803

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System, 28*(4), 499-509. Retrieved from http://dx.doi.org/10.1016/S0346-251X(00)00035-X

Wall, D. (2005). *The impact of high-stakes examinations on classroom teaching: a case study using insights from testing and innovation theory*. Cambridge: Cambridge University Press.

Waters, J. K. (2010). Enter the iPad (or not?). *T.H.E. Journal, 37*(6), 38-40.

Weir. C. J. (2005). *Language testing and validation: an evidence-based approach*. New York, NY: MacMillan.

Whittaker, R. (2006). Review of Estudios y Criterios para una Selectividad de calidad en el examen de inglés. *Estudios Ingleses de la Universidad Complutense, 14*, 198-206.

Research-publishing.net

New perspectives on teaching and working with languages in the digital era
Edited by Antonio Pareja-Lora, Cristina Calle-Martínez, Pilar Rodríguez-Arancón