Preliminary Findings of the PAULEX Project: A Proposal for the Internet-based Valencian University Entrance Examination

Jesús García Laborda
University of Alcalá/Modern Languages, Madrid, Spain
Email: jesus.garcialaborda@uah.es

Abstract—Although there is an increasing number of online tests in the world, little research is currently known in Spain today. Assessment has become an integral part of education and the implications of the various uses of language testing go beyond the educational settings (Douglas, 2010; Fulcher, 2010). This study describes the PAULEX project. This project deals with the design and validation of a computer-based University Entrance examination for Spanish universities. The test measures the students' oral and written competence including tasks on four different skills: reading, writing, listening and speaking. This paper will address some specific features and reaction studies that are already in process. It is by no means this paper's object to give detailed information of each individual aspect but instead to give an overview of what has been done so far and provide concepts of the current research. In doing so, the paper is divided into three main sections: technological advances, attitudinal observations and results.

Index Terms—testing, English as foreign language, computers, attitudes

I. Introduction

Although there is an increasing number of online tests in the world, little research is currently known in Spain today. Assessment has become an integral part of education and the implications of the various uses of language testing go beyond the educational settings (Douglas, 2010; Fulcher, 2010). The reasons could be that computers require economic efforts from the researching institution at the beginning and that researching on new field requires specialists of more areas than those demanded for traditional testing (Bachman, 2000; Hambleton & Slater, 1997). Probably, these are also the reasons why the Spanish Ministry of Education may provide funding for research projects in the field, but the implementation of those projects may, in reality, be far for being put into practice. The PAULEX project, whose findings are hereby succinctly presented, may have contributed to the general knowledge in the area but most likely will take a long time (if ever) to be considered worth putting into practice. This article briefly reviews the findings of the PAULEX project and suggests reasons to make the web-based University Entrance Examination (IB PAU) real.

II. Computer-based Language Testing

Computer-based language testing has been with us since the mid-eighties, as a way to deliver language tests via the computer. According to their construct, they can be closer to traditional or adaptive tests (Economides & Roupas, 2008; Ockey, 2009; Yen *et al.*, 2010) based on item-response theory (Colpaert, 2004) in which providing the correct response to an item leads to a more difficult one and vice-versa. When compared to traditional tests, Computer Assisted Language Testing (henceforth CALT) offers a large number of advantages, such as the inclusion of visual media, the delivery of spoken items without the need of human resources (Ockey, 2009; Siozos *et al.*, 2009; Tung & Deng, 2006), a larger type of integrated and analytical test items, or rapid feedback (immediate when dealing with objective items). Additionally, online delivery permits students to take the tests at lower costs from their schools. However, the most important effects are based on the validity of the test, especially because the test conditions (with very few exceptions)[1] are the same for all the candidates (Pardo-Ballester, 2010). Until recently, Roever (2001) asserts that preparing those tests was more difficult than traditional ones and also more expensive to produce and maintain. However, recent advancements in web-based tests have facilitated the maintenance operations and reduced the number of hours, since the same problems may be solved for the whole system instead of checking each individual computer. Besides, each computer does not need to have the whole software installed, but may work with just a number of plug-ins. At present, most of the important computer-based tests use a combination of traditional and self-corrected items. For instance, TOEFL has an integrated skills system (García Laborda, 2006; Meskill & Rangelova, 1995) that allows a chain of tasks corresponding to different skills such as speaking-listening-reading-writing (with all the combinative possibilities).

---

[1] In high stakes tests, these exceptions are usually well known, such as the security breakthrough in the Turkish OSYM in 2010 (http://www.todayszaman.com/news-222239-many-security-gaps-found-in-osym-as-police-arrest-four.html) or the TOEFL Internet cut during a test in Shanghai, Beijing, Guangzhou and Hangzhou (http://www.guardian.co.uk/education/2006/oct/20/tefl2?INTCMP=SRCH).

Likewise, the PLEVALEX and PAUER platforms share many features with the TOEFL, although the Spanish platforms carry out skills integration in a less sophisticated and more linear fashion.

III. How Was the Spanish PAULEX Designed?

This project dates back to the original HIEO/HIELE project started in the Universitat Politecnica de Valencia (UPV), where a group of experts noticed the need to have a more efficient testing system to diagnose international students arriving at that university. The system consisted of five main modules, including grammar, reading, writing, listening and speaking (García Laborda, 2006). The testing delivery system was designed by the experts at the UPV and implemented by a company in Madrid. The experimentation was rather limited, but100 French or English students took a test through that system. The results of the positive experience were sent to the Generalitat Valenciana. A second version of the platform was implemented only for the writing module. The section, although a bit limited, was implemented and later incorporated to the PLEVALEX testing platform. PLEVALEX was the second and improved design of HIEO. After further experimentation, in 2007 the current PAULEX/PAUER platform was implemented. This platform began to be operational in 2009 and tests of English and Spanish were administered. In fact, by mid 2010, more than 300 students had taken an Internet-based version of the English or French PAU, both in the university campus (and net) system and outside it (Colegio El Pilar and Instituto Benllure of Valencia). The platform had also served to train a group of 30 teachers on the integration of computers in the PAU. By the end of 2010, it was clear that the IB PAU could be implemented and that students would not reject the testing delivery system. Furthermore, studies were conducted to observe their reactions towards the implementation of the PAU through mobile phones and devices with promising results.

IV. What Did We Study in These Experiences?

As mentioned in the previous section, overall the system combined many research aspects. The research team was mainly divided into two sections: language testing and computer/software development. Both teams have spent the last years providing an immense quantity of information.

A. Information on Language Testing

This team varied in size and a number of experts took part in the different aspects of the design and implementation stages. This team worked alongside computer engineers who designed the interfaces. Within the three areas that were the main intended research subjects, we focused on technology development and its effects in humans. However, previous stages took us to other sub-studies. These studies were necessary to focus the research adequately:
- Desire and need to change the University Entrance examination.
- Current problems of high school students in their foreign language classes.
- Advantages and disadvantages of high stakes testing.
- Attitudes towards the new oral tasks.

B. Information on Computers and Their Effect on the Stakeholders

Obviously, the design of a computer-based test requires a deep study of the interaction between the candidate (human user) and the computer. Much of the interaction in the testing situation occurs at the testee interface level (Fulcher, 2003). Thus, there are two main elements that participate in the final design of the delivery system, which are the hardware and the software. The PAULEX project technological development of interfaces was based on the results suggested by García Laborda (2006), which lead to an interface design similar to TOEFL (García Laborda, 2006) and quite different from other tests like IELTS or BULLATS (García Laborda, 2007). On the other hand, the software principles and database were based on the principles set by Gimeno Sanz in 2002 and used for the Ingenio language teaching and learning platform. As for the PAULEX project, the following are the fundamental design principles that were addressed in different projects:
- Previous interface design, especially Flucher's principles (2003).
- A comparison of different computer-based tests.
- Effect of image and computer context in the candidates.

V. Experimental Phase

A. The Test and Its Implications

The language testing experimentation began through collecting the teachers' attitudes towards the still current PAU and a new one based on the introduction of speaking and listening tasks. García Laborda & Fernández Álvarez (2010) discovered that teachers were eager to implement a dramatic change, although they also feared the effects of such a change.

We also found that many teachers were excited about the impact of technology, but many felt overwhelmed by the excess of responsibility of getting high grades in the PAU. All these fears and anxiety were, naturally, due to the

possibility to include the speaking tasks (listening did not have such impact on the teachers' attitudes), which most teachers thought to be based on short conversations in pairs and short descriptive monologues.

In relation to the construction of the test, we found evidence that supports that the change would have an impact almost uniquely on the speaking tasks. Their results supported the fact that, in the long run, due to the washback effect (Amengual Pizarro, 2009; García Laborda, 2008; Green, 2007; Hausenberg, 2006; Luxia, 2007; Watanabe, 2004), students would improve their speaking ability. They also stated that converting current PAU tasks into computer tasks would have limited differences. The last part of the studies conducted in relation to language testing was done in teacher training. In order to foresee future effects of the new test, a group of 27 teachers was trained in Valencia in a teacher development course for continuing professional development (CPD). These in-service teachers took a 30-hour course that included both the development of their capacities as testers and training in the use of technology for education. Results indicated that when appropriate training is provided and teachers know when, why and how to test, the effect of technology diminishes (García Laborda & Litzler, 2011).

B. Looking at Computers

The results were obtained through experimentation of software and, above all, interface information. In some papers, the research group found that students would be able to adapt to the new test delivery system (Figs. 1 and 2).

Figure 1. High school student takes the IB PAU (2008 version).

Figure 2. Testing conditions in the IB PAU (2008 version).

As can be seen, the development of the testing interface (Fig. 3), the buttons and the navigation elements were reduced and also followed the design principles stated by Fulcher in 2003. This interface design has been informed in a number of papers (García Laborda, 2009).
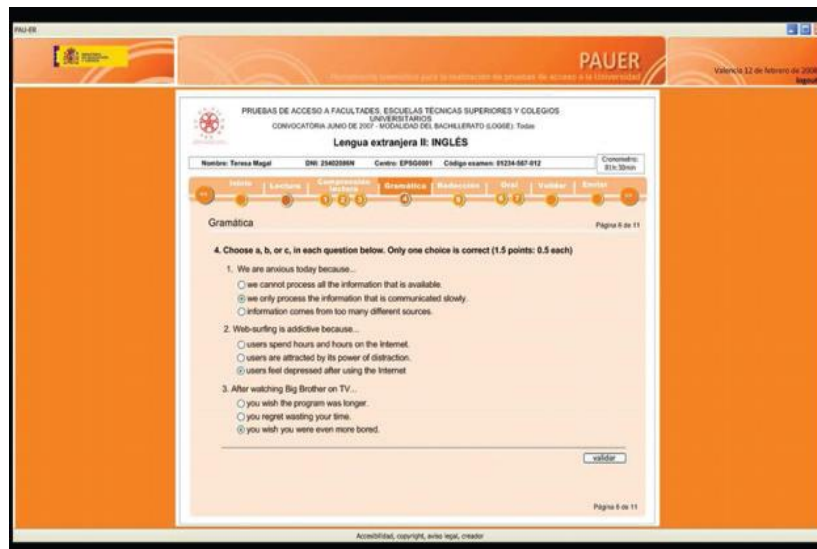
Figure 3. PAULEX interface design.

The most significant advances, however, were made in relation to the students' adaptation to the computer-delivered language testing. A number of studies evinced that, although students may face the test for the first time, they can adapt easily to most tasks. Overall, the ones to which students adapt more easily are those that can be responded with a tick or response selection (such as multiple choice tasks). Field studies also found that, despite the teachers' attitudes, most difficulties were associated to writing. Other aspects of research included the use of mobiles for testing through a testing emulator (Fig. 4).

Figure 4. Mobile testing through the use of emulators.

VI. DISCUSSION AND IMPLICATIONS FOR THE CLASSROOM

The final goal of this project was to obtain evidence of the processes and contexts in which the onlining of the University Entrance Examination (PAU) could be feasible. This research was predominantly practical. Practicality, which was not considered initially, came across as a valuable asset because the crisis in the last two years also requires the limitation of test production and delivery costs. Additionally, practicality was considered because educational boards may also consider the costs of a face to face speaking test done either in the students' high school or in the universities location.

Testing implications

Computer tests should reduce costs but also improve the construct validity in relation to formats and constructs that are used today. However, although the four different skills to be assessed would be interrelated, they also need to be analyzed through factor analysis and intersection correlation when the official trial tests begin (if ever).

VII. CONCLUSIONS

Our project and its related studies showed that it is possible and reliable to implement an Internet-based PAU. However, this study can only be considered a first approach and the final validation may depend to a large extent on the regional educational testing boards' interest in trialing, redesigning and implementing the testing platform and test construct in a near future. Thus, a genuine belief in the tremendous possibilities of this test delivery system is indispensable. That means that much work still needs to be done. The project, the most ambitious one in Spain to this date, requires both a technical and research continuation that in this period of national crisis seems too difficult to achieve. Testing advances require an expert definition of the construct with the integration of different items and skills. Since developing automatic scores in writing and, especially, speaking seems too distant for now, professional training is necessary (Amengual Pizarro, 2006). Finally, the limited number of projects of this type in Spain, however, may be a reason to continue research in this field.

REFERENCES

[1] Amengual Pizarro, M. (2006). Análisis de la prueba de inglés de selectividad de la Universitat de les Illes Balears. *Iberica* 11, 29-59.
[2] Amengual-Pizarro, M. (2009). Does the English test in the Spanish university entrance examination influence the teaching of English? *English Studies* 90.5, 582-598.
[3] Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing* 17.1, 1-42. doi:10.1191/026553200675041464.
[4] Colpaert, J. (2004). Editorial: Transdisciplinarity. *Computer Assisted Language Learning* 17.5, 459-472.
[5] Douglas, D. (2010). Understanding language testing. New York: Oxford University Press.
[6] Economides, A. A. & C. Roupas. (2007). Evaluation of computer adaptive testing systems. *International Journal of Web-Based Learning and Teaching Technologies* 2.1, 70-87.
[7] Fulcher, G. (2003). Interface design in computer-based language testing. *Language Testing* 20.4, 384-408.
[8] Fulcher, G. (2010). Practical language testing. Great Britain: Hodder Education.
[9] Garcia Laborda, J. (2006). International testing platforms: When similar outcomes meet different testing needs. Paper presented at the 4th International Conference on Multimedia and Information & Communication Technologies in Education (m-ICTE2006) (22-25 Nov., Seville, Spain). http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED504792&ERICExtSearch_SearchType_0=no&accno=ED504792 (accessed 1/10/2011).
[10] García Laborda, J. (2006). La plataforma de exámenes multilingüe PLEVALEX: Resultados del diseño y perspectivas de investigación futura de la plataforma de exámenes valencia de lenguas extranjeras. *Didáctica: Lengua y Literatura* 18, 135-145.
[11] García Laborda, J. (2007). Introducing standardized EFL/ESL exams. *Language Learning & Technology* 11.2, 3-9. http://llt.msu.edu/vol11num2/pdf/net.pdf (accessed 1/10/2011).
[12] García Laborda, J. (2008). Enhancing washback projections of the PLEVALEX online testing platforms. In N. Kunt, J. Shibliyev & F. Erozan (eds.), *ELT profession: Challenges & prospects*. Germany: Lincom Europa. 129-126.
[13] García Laborda, 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 Psicopedagogía*, 21.1, 71-80.
[14] García Laborda, J. & M. Fernández Alvarez (2010). Las variables sexo, edad y lugar de trabajo en las actitudes de los profesores hacia la interacción oral en L1 y L2 en la clase de inglés de segundo bachillerato. *Porta Linguarum* 14, 91-103.
[15] García Laborda, J. & M. F. Litzler (2011). Constraints in teacher training for computer assisted language testing implementation. *International Journal of Education Studies* 4.2, 13-17.
[16] García Laborda, J., M. López Santiago & A. Martínez Sanz (in press). French language adaptation to CALL testing environments. In A. Gimeno Sanz (ed.), *New trends in CALL: Working together*. London: Macmillan.
[17] Garcia Laborda, J., T. Magal Royo & M. Bakieva (2010). A first approach to the analysis of student motivation in the trial version of the computer based university entrance examination. Paper presented at the 14th CALL conference (18-20 Aug., Antwerp, Belgium).
[18] Gimenez López, J. L., T. Magal-Royo, F. Garde Calvo & S. Prefasi Gomar (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.
[19] Gimeno Sanz, A. (2002). Call software design and implementation: The template approach. Valencia: Servicio de Publicaciones de la Universidad Politécnica de Valencia.
[20] Green, A. (2006). Watching for washback: Observing the influence of the international English language testing system academic writing test in the classroom. *Language Assessment Quarterly* 3.4, 333-368.
[21] Hambleton, R. K. & S. C. Slater (1997). Item response theory models and testing practices: Current international status and future directions. *European Journal of Psychological Assessment* 13.1, 21-28. doi:10.1027/1015-5759.13.1.21.
[22] Hausenberg, A. (2006). Washback of testing on language learning and teaching [Eksamite tagasimoju keeleoppele]. *Eesti Rakenduslingvistika Uhingu Aastaraamat* 2, 43-52.
[23] Luxia, Q. (2007). Is testing an efficient agent for pedagogical change? Examining the intended washback of the writing task in a high-stakes English test in China. *Assessment in Education: Principles, Policy & Practice* 14.1, 51-74.
[24] Meskill, C. & K. Rangelova (1995). US language through literature: A transatlantic research project. In M. Warschauer (ed.), *Virtual connections: Online activities and projects for networking language learners*. Honolulu, HI: University of Hawai'i Second Language Teaching and Curriculum Center.

[25] Ockey, G. J. (2009). Developments and challenges in the use of computer-based testing for assessing second language ability. *Modern Language Journal* 93.S1, 836-847. doi:10.1111/j.1540-4781.2009.00976.x.

[26] Pardo-Ballester, C. (2010). The validity argument of a web-based Spanish listening exam: Test usefulness evaluation. *Language Assessment Quarterly* 7.2, 137-159.

[27] Roever, C. (2001). Web-based language testing. *Language Learning & Technology* 5.2, 84-94.

[28] Siozos, P., G. Palaigeorgiou, G. Triantafyllakos, & T. Despotakis (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. doi:10.1016/j.compedu.2008.12.006.

[29] Tung, F. & Y. Deng. (2006). Designing social presence in e-learning environments: Testing the effect of interactivity on children. *Interactive Learning Environments* 14.3, 251-264. doi:10.1080/10494820600924750.

[30] Yen, Y., R. Ho, L. Chen, K. Chou & Y. Chen. (2010). Development and evaluation of a confidence-weighting computerized adaptive testing. *Educational Technology & Society* 13.3, 163-176.

[31] Watanabe, Y. (2004). Methodology in washback studies. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Jesús García Laborda, PhD, EdD, MA, Med, BA, is currently an associate professor at the Universidad de Alcalá (Madrid, Spain). He received both his PhD in Linguistics (2000) and his EdD in Language Education from the Universidad Complutense de Madrid. His MA was obtained from the University of Wisconsin-Milwaukee (1993) and his Med from the University of Georgia-Athens (1992). Besides his focus on TESOL, his recent publications include language for tourism, teacher training and education. He has published in some international journals such as *Computers & Education*, *Educational Technology & Society*, *Eurasian Journal of Educational Research*, *Modern English Teacher* and many more, with a total of over 200 academic papers. He has taught in Spain, the United States and in Bosnia & Herzegovina and Croatia during their reconstruction after the wars in the late 90's. His main areas of interest are: (1) computer assisted language learning, (2) ESP, (3) low stakes testing, and (4) teacher training.