21 From monolingual to bilingual through testing: The OPENPAU Project (FFI2011-22442)

Jesús García Laborda
Mary Frances Litzler
Universidad de Alcalá

Abstract

Teachers in Spain have long criticised the effect of language testing in their classrooms. For example, García Laborda and Fernández Álvarez (2012) discuss the reactions of teachers from Navarre (northern Spain) and Madrid towards the country's current University Entrance Examination (*Prueba de Acceso a la Universidad*, henceforth, PAU), the most important high-stakes test in the country. The exam has no speaking or listening tasks, and the teachers feel that this format impedes development of oral skills. Although the potentially positive impact of including speaking and listening tasks (Alderson and Wall 1993, Hughes 1989) is one of the main reasons for changing the PAU, the delivery and rubrics of the components must first be designed (García Laborda and Martín Monje 2013). The main goal of this paper is to address what has been done in the last year in relation to these issues. Some of the results of research carried out through the OPENPAU project in Spain, which emphasises the application of technology to individual and paired/group testing, are described here. The paper concludes that the use of low cost technology for the test may have a very positive impact on the Spanish educational system. It also suggests that the integration of speaking tasks at the end of different educational cycles can also have a potentially positive washback effect (Wall 2000, Wall and Alderson 1993).

Introduction

As in many other countries around the world, people regard English highly in Spain for its value in commerce, media and work. Administrators, parents and students feel that one of the most important outcomes required by the end of secondary school is mastery of English (Breeze and Roothooft 2014).

452

The level of foreign language proficiency required changes from one educational region to another, but a general aim in Spain is for students to achieve the level of B2 on the Common European Framework of Reference (CEFR, Council of Europe 2001) upon completion of secondary school (Ashton, Salamoura and Diaz 2012).

Nevertheless, the results of international surveys show that, despite intense efforts made by the different national and regional educational administrations, Spanish students lag behind most other European countries in knowledge and use of English as a Foreign Language. For instance, according to the European Survey of Language Competences (European Commission 2012), Spanish students begin to learn foreign languages at the age of six, when they begin their compulsory education, but at the age of 15 as many as 32% of the students have only achieved a pre-A1 level in listening comprehension and 15% show this level in written expression (Figure 1).

These data may reflect the limited attention that listening skills receive in the classroom for students through the age of 15 (the end of compulsory secondary education in Spain) (García Laborda and Fernández Álvarez 2011). However, additional factors may also have a negative effect on the

Figure 1 Percentage of students achieving broad CEFR levels by skill and country/educational system (First target language)

Educational system	Language	Reading			Listening			Writing		
		Pre-A1	A	B	Pre-A1	A	B	Pre-A1	A	B
Bulgaria	English	23	43	34	23	37	40	15	52	32
Croatia	English	16	44	40	12	32	56	5	49	45
Estonia	English	7	33	60	10	27	63	3	37	60
Flemish Community of Belgium	French	12	63	24	17	62	20	19	59	22
France	English	28	59	13	41	46	14	24	61	16
French Community of Belgium	English	10	59	31	18	55	27	6	65	29
German Community of Belgium	French	10	52	38	11	49	40	8	51	41
Greece	English	15	40	45	19	35	46	7	41	53
Malta	English	4	17	79	3	11	86	0	17	83
Netherlands	English	4	36	60	3	21	77	0	39	60
Poland	English	27	49	24	27	45	28	19	59	23
Portugal	English	20	53	26	23	39	38	18	55	27
Slovenia	English	12	42	47	5	28	67	1	51	48
Spain	English	18	53	29	32	44	24	15	58	27
Sweden	English	1	18	81	1	9	91	0	24	75

Source: First European Survey on Language Competences (Surveylang 2012:92)

453

learning of the language, such as the correlation between literacy in the L1 and L2 (Ganschow 1991), overuse of the first language in the classroom (Atkinson 1987), differences between the mother tongue and the foreign language (Chang 2009, Hutzler, Ziegler, Perry, Wimmer and Zorzi 2004), and others that make the learning of English harder for Romance language speakers than for other Europeans whose languages have a Germanic origin. While these issues may justify in part the fact that Spanish students find learning English problematic, the objective of the European Union for its citizens to speak at least two foreign languages by 2020 requires an urgent and robust solution.

Testing as a catalyst for change

There has been extensive debate over the impact of testing on the capacity to trigger learning (Green 2007, Nemati 2003, Perrone 2011, Spratt 2005) and on teachers' attitudes to teaching methodologies and materials (Cheng 1997, 1999, Turner 2006), both of which areas fall under the concept of 'washback' (Alderson and Wall 1993). García Laborda and Fernández Álvarez (2012) asked teachers when they felt their schools should begin specific instruction for speaking tasks if the PAU were to begin including them. The study revealed that some teachers thought that preparation should begin up to four years in advance. The research team was perplexed at this amount of time that would be devoted specifically to preparation for test tasks as opposed to opportunities for genuine learning.

If the test really were to have such a powerful influence, could the recent introduction of external English exams which include speaking tasks possibly trigger the introduction or reinforcement of speaking skills by most teachers and thus provide an indication of the effect of introducing a speaking component in PAU? External exams are currently being used in the relatively new state bilingual schools in the Madrid region (Dobson, Pérez Murillo and Johnston 2010). The programme is especially strong at the primary level and consists of 9 to 11 hours per week of instruction in content (typically maths and science) and language classes taught in the target language, which is often English. One of the most important activities to promote this programme is success on external language tests at the end of the sixth year. In general, schools prefer that their students take either Trinity or Cambridge English tests. Although the idea is motivating for students, it is even more so for parents, who are happy to see their children pass official tests. In fact, many parents seem to feel the level of an official exam is less important than the grade or mark obtained. In other words, a 'pass' on the *Cambridge English: Preliminary* (also known as Preliminary English Test (PET)), a test at the B1 level of the CEFR, is considered much the same as 'a pass' on the *Cambridge English: Key* (also known as Key English Test (KET)), at the A2 level, and

a high mark in a lower proficiency level may even be viewed in a better light than a lower one on a more demanding test. Nevertheless, the bilingual programme is not without its problems, for instance, some students may not be able to adapt to a bilingual situation, especially in maths and science and may decide to change to regular non-bilingual programmes (Anghel, Cabrales and Carro 2012).

The Spanish Ministerio de Educación (2013, 2014) and other researchers (Spratt 2005) suggest that tests taken during the primary school period, especially external ones, may have a positive effect in developing positive washback. That is, they seem to impact classroom practice for the better, first by changing teachers' instructional practices and second by increasing motivation for achievement. However, more work is needed to determine how tests can potentially foster language acquisition at the secondary level. Possibilities for researching this level will increase as the bilingual programme is expanded in the coming years.

Towards an adequate testing procedure: The OPENPAU project work in progress

The OPENPAU project was funded by the Spanish Ministry of Education in 2010 (it has been administered by the Ministry of Science and Innovation from 2011). Its main goal is to obtain information on how to improve language results through the testing of 15 to 17-year old students in post-secondary education in Spain and to suggest ways to improve the current situation. One of the most important parts is the design of a technology delivered test.

In 2010, 22 researchers from ten different universities in Spain and another from the US participating in the OPENPAU project were assigned different tasks, including a needs analysis, an analysis of the current situation in English classrooms, and a review of technology for exam delivery. A group of psychometricians also participated when required. The major outcomes of the needs analysis were detailed in an earlier paper (García Laborda and Martín Monje 2013) and the findings of the second phase of work were presented at the ALTE conference in Paris. This section will address two main issues: the current situation in English classrooms in Spain and three proposals for delivery of the test.

The English language classroom in Spain today

In order to investigate the current situation in English classrooms in Spain, the following activities were undertaken. Forty native speakers of English working in the Madrid region as conversation assistants who were also studying towards a Masters degree in Bilingual Education or in

455

International Education were asked to report on their experience in schools. They reported their impressions as part of a teaching portfolio completed in partial fulfilment of the Masters' degree programme. The two cohorts of 20 subjects each (one for each Masters programme) provided the following information for the Madrid area. Additional information on the rest of Spain was obtained from 27 coordinators of the University Entrance Exam who completed an online survey to point out problems detected in the exam. The results revealed by these different groups are summarised below.

Spain's foreign language classrooms have seen the arrival of an increasing number of native English speakers as conversation assistants, most of whom are assigned to the primary school level. In general, they lack language teaching skills and their previous experience is very limited, as the only educational requirement is to have a university degree. In addition, they often have limited knowledge of linguistics and language learning processes, and are only provided with an orientation of a few days' duration before starting at their schools. Interaction between conversation assistants and individual classes of students can be limited to 1 hour a week per group as some schools have only one conversation assistant for the entire school, which may have as many as 600 students. The target language speakers often support the teaching of Social and Earth Sciences and English language but they are not limited to doing so and sometimes they have to support teachers in subjects in which they have not specialised. Given the situation described in general here, there is often a gap between their preparation and their responsibilities, what they are expected to do and what they have been trained to do.

A second issue that was observed is that there is limited stimulation of actual language use through interaction due to discipline problems in the classroom. Many teachers must focus more attention on controlling their students' behaviour than on practising the language. Students have a tendency to talk amongst themselves and can even get up out of their seats during classroom explanations so many teachers find that they frequently need to call their groups to order. In addition, it is not unusual for teachers of other subjects to complain that there is 'too much noise in the school', so teachers tend to avoid extensive speaking activities in class. Student claims that their classes are taught in Spanish were also common, but the research team felt that this situation was becoming more unusual.

Related to this issue is the avoidance of pair and group exercises due to a lack of control of what goes on in the different groups around the classroom. With as many as 35 students in a classroom, it is impossible to monitor the different groups at all times. This kind of interaction is also avoided to assess speaking since most teachers would opt to do individual tests with their students, but their time is limited and they are unable to do so.

Another issue that has been seen is the discrepancy between what teachers assess and what students actually do in class. While classes do tend to be

456

progressively communicative, teachers still often follow a textbook and workbook in their day-to-day lessons due to pressures to complete them from administrators and parents who have paid for them. Teachers in both private and public schools still often overuse grammar or discrete point tests. On many occasions, teachers were observed to use textbook tests despite incorporating more communicative activities in their classrooms.

Due to these factors students over-regulate their language production on tests as well as in real life situations. In general, they feel uncomfortable and are unsure about their ability to apply correct grammar and vocabulary and to pronounce the language well in spoken interaction. They also demonstrate a high level of anxiety about external exams.

In terms of test performance and behaviour, a sub-study within this project revealed that the PAU coordinators observed in internal reports student difficulties in doing the PAU. When they write, students tend not to read the rubric properly, they interpret the wording of the questions literally, are unable to recognise the text type, cannot identify appropriate introductions or conclusions, and do not know how to organise their essays. The research team of this project also found in a mock-PAU speaking test, that the students tended to translate literally, did not parse what they were saying, and showed limited use of vocabulary to describe pictures. They also had difficulties in turn taking, showing interest, and expressing disagreement, etc. (i.e. communicative functions). At the same time, they did not listen actively to the examiner or their partner, and they lacked adequate content and form in their responses.

A Delphi study to improve language learning through testing

In order to improve this situation, a group of nine experts were asked to do a Delphi analysis to provide proposals to improve foreign language learning in Spain through testing. The Delphi Method is a systematic and interactive method in which a panel of experts answer a questionnaire in several rounds (generally between two and three) based on the other experts' responses and opinions on the same questionnaire. As a result, they must read the other participants' replies each time before answering the following time (Rowe and Wright 1999). In this case, the researchers suggested a list of questions, the nine experts responded individually and then their responses were sent to all the other experts. Afterwards, an online video conference was held. The original questions posed were as follows:

- How can tests benefit Spanish students' development of language skills?
- How can tests help students to improve their speaking/writing skills?
- What activities should the research group do to carry out the project?

The experts' individual comments and suggestions led to the following outcomes:

- compulsory speaking tests should be introduced (García Laborda and Fernández Álvarez 2012)
- specific in-class preparation and training for testing skills should be reduced as previous experience has shown highly negative washback involving considerable emphasis on grammar
- teacher training for speaking must be improved (Fernández Álvarez and García Laborda 2011, García Laborda 2013)
- a speaking construct and standards for performance required to achieve the ideal level should be defined (García Laborda 2010)
- consideration should be given to moving away from the current cognitivist paradigm in which students are asked to provide evidence of language knowledge rather than a capacity to communicate in the foreign language (Bejarano, Otero de Juan and García Laborda 2013, Canagarajah 2006)
- a social-constructivist/interactionist approach should be introduced and it should specifically emphasise interaction and mutual support between/among test takers
- the ICT tools to be used should be decided (García Laborda, Magal Royo, Litzler and Giménez López 2014).

The online meeting led to the following conclusions, which were later forwarded to the Ministry of Education. Specifically, the test construct, impact, delivery and testing sessions need to be revised. The test construct should include three speaking and two listening tasks in addition to the already existing exam format. The three speaking tasks represent what the experts considered the most important aspects of a student's use of a foreign language: dialogues about personal issues; presentations followed by questions, and organisation acts such as preparing for a meeting or exam, etc. The listening tasks should address monologue comprehension (i.e. a teacher's class presentation) and a social situation. These tasks are in line with Bygate (1999) and Weir (1993), who consider them informational and interactional functions.

It was also emphasised that there should be a continuous push towards speaking in class and the inclusion of speaking in tests starting in the primary level in order to increase their impact on learning. The tests should be required at the end of the second, fourth, sixth, tenth and twelfth years. Nevertheless, teaching for the tests should be avoided. In terms of delivery, the speaking tasks for the PAU should be done in pairs by two examiners. This could be done in person or online through the use of tablet PCs or netbooks. The speaking interviews should be limited to pair tests lasting a

maximum of 15 minutes. If the delivery is done online, the productive time should be about 5 to 6 minutes per person.

A SWOT analysis for a speaking test for the PAU

A SWOT analysis of the potential inclusion of a speaking test as part of the PAU was also completed in this analysis phase of the project (Table 1):

Table 1 presents the SWOT analysis drawn up by several participants in the Delphi analysis, three external experts and the research team. The strengths are related to the potential to meet the demands of both teachers and students to create and assess speaking skills in the PAU (and other tests) and the adequacy of the test construct for the purpose (face validity), washback (impact validity) and delivery. The weaknesses include a lack of teaching experience in teaching for a new test (as in the case of Saif 2006), which leads to a need for training and specific preparation in technology both for teachers and for students (Akinwamide 2012). The new test would have a positive effect on curriculum changes introduced as a result of a new educational law recently passed in Spain, and it would lead to new ways of teaching and learning for teachers and better learning and external certification of knowledge for students. These opportunities could potentially be jeopardised by problems in learning required for the test, abuse of teaching for the test, stakeholder difficulties in understanding the benefits of external and internal assessments (the culture of testing), student unwillingness to adapt to technology delivered tests or to group tests, the overuse or excessive potential of tests to affect decisions (negative washback), and the difficulties intrinsic to the increasing importance of implementation of new skills in the process in foreign language learning.

Proposals for delivery

One of the most significant issues in planning the future University Entrance Examination is delivery (Figure 2). As mentioned above, one idea is to have

Table 1 SWOT analysis

Strengths:	Weaknesses:
1. Teachers' interest in improving their classes	1. Teaching resources
2. Students' motivation to improve	2. Teaching methods
3. Testing construct and FL learning	3. Teacher training
4. Ways of delivery	4. ICT issues
Opportunities:	**Threats:**
1. Policy-makers: Accountancy	1. Cultural perspective
2. Teachers' visions of language improvement	2. FL difficulties
3. Students' perceptions of the importance of English	3. Students' volition towards the new testing methods and delivery
	4. Overpowering tests (negative washback)

Figure 2 Delivery modes of PAU test
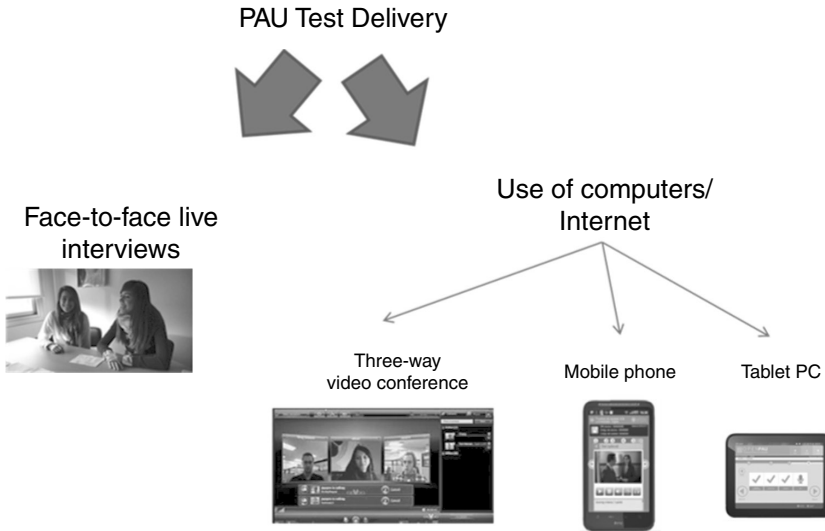
group interviews in face-to-face situations while another is to integrate computer and Internet technologies in the test. While the latter suggestion relates to the use of tablet PCs and mobile phones, both possibilities could be done through the use of Voice-over-Internet-Protocol (VOIP) programmes such as Skype, Tango or Viber.

While the first two of these computer applications, tablet PCs and mobile phones, are viewed by the research team to be more efficient in terms of practicality, facilitate and trigger individual production, videoconferencing could be optimal if factors such as support, body language, language interaction and others are taken into account; all of these aspects of communication provide a more complete view of the speaker's communicative competence in oral situations. In terms of organisation, the first two applications are considered slightly better than the third but the research team also thinks that organisation of a VOIP video conference should not be problematic in high-stakes testing if Ministry secured lines are used. It is also recommended that students should be supervised by their own teachers or that other security methods – such as outside invigilators – should be set up to avoid potential cheating. The cost of delivery was another aspect also taken into consideration by the research team. Technological applications involve a design cost and mobile testing requires hardware, but VOIP technology is almost free even if a subscription rate for private channels is paid. For example, Skype allows videoconferencing from three or more computers for free, so two test takers and the examiner(s) would be able to see each other during the test session. In terms of internet security,

mobile phones and tablets both require strong protection against any kind of attack, while in the case of VOIP many of the measures used to protect online platforms could be simplified because the test items are delivered in real time.

Conclusion

In conclusion, the research team members believe that testing can lead to improvement in language learning as reported for the Sri Lanka project in Wall (2006), who mentions improved teaching methodologies, development of the neglected skills of speaking and listening, a revision of the legal mechanisms to improve teaching quality, a revision of classrooms practices, and so on. The researchers also feel that the use of mobile ICT technology can foster ongoing practice in and out of the classroom, as described in Towndrow and Wan (2012) and Chen, Chang and Wang (2008). At the same time, the methods suggested in the present paper can facilitate assessment of a large number of candidates by using paired interviews (or 3-person groupings) (Valk, Rashidand and Elder 2010). In addition, the continuous practice of speaking needed to achieve a level potentially required to pass a test can lead to acquisition of the language (Muñoz and Álvarez 2010). As a consequence, teachers may change their practices and methods (Spratt 2005).

Future steps in the OPENPAU project include the design of new types of collaborative tasks for all four skills, in particular, cooperative writing, and the application of social networking to language testing. The project will finish by the end of 2015. The changes recommended by the research team may have political implications, such as better indicators and information for educational policy-makers and teachers and increased European labour integration, as suggested in Prapphal (2008). In conclusion, the introduction of language tests in general education, especially in post-compulsory education, can benefit the whole system and its different stakeholders, be they students, teachers or administrators. We feel, as the title of this paper indicates, that tests with communicative oral tasks delivered using new technologies will lead to a higher level of bilingualism throughout the country in the coming years. As oral tests are included throughout the different years of compulsory education, a greater emphasis will be placed on oral skills in the classroom. The end result will be better productive ability on the part of Spaniards compared to today.

Acknowledgements

461