ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 4, No. 9, pp. 1763-1769, September 2014
© 2014 ACADEMY PUBLISHER Manufactured in Finland.
doi:10.4304/tpls.4.9.1763-1769

Setting the Standards for the Foreign Language Speaking Tasks of the New Baccalaureate General Test

Mª Camino Bueno-Alastuey
Public University of Navarre/Philology and Didactics of Languages, Pamplona, Spain

Jesús García Laborda
University of Alcalá/ Modern Philology, Alcalá, Spain

Ana Isabel Muñoz Alcón
Avila University/Avila, Spain

Gloria Luque Agullo
Jaen University/ English Philology, Jaén, Spain

Abstract—**One of the most significant aspects of the Spanish new educational reform is the Baccalaureate General Test which is intended to replace the former University Entrance Examination. The new test will include an oral part, which needs to be created, based on current research on the field (Bueno-Alastuey & Luque, 2010), and tested with students from different regions in Spain to confirm its validity. This paper describes the preparation and first results of a pilot study using some proposed tasks. The speaking tasks were based on the ones currently used in the Cambridge Preliminary English Test but conveniently adapted to the Spanish context as suggested by some studies (Amengual-Pizarro & Mendez García, 2010). This paper shows the perceived strengths and weaknesses of those tasks and the test based on current testing literature on construct definition (Bachman & Palmer, 1996) and validation (Weir, 2005; Fulcher, 2010; Ekbatani, 2011). Results showed that the test corresponds better to classroom practice and favors both washback and language development at a lower cost.**

Index Terms—**Spain Baccalaurate General Test, speaking competence testing, pilot speaking test, oral test validation**

I. Introduction

One of the most significant aspects of the Spanish new educational reform has been the introduction of the Baccaulerate General Test, which will replace the former University Entrance Examination. Apart from the fact that this test is to be carried out in a different setting (the high school where each students is enrolled instead of at the university), another major change is that the test will include an oral component to test receptive and productive oral communicative competence. The different setting may help to reduce the anxiety this kind of high-stakes tests produce because the setting will be more familiar and, thus less threatening, and it will imply a reduced number of students taking the test at the same time in the same place. Both facts will probably provide a more supportive environment for the students. However, the inclusion of an oral competence part will greatly increase anxiety as oral skills are the ones that cause more anxiety in Spanish students.

Previous research has put forward proposals for the tasks to be included in the oral part of the test (Bueno-Alastuey & Luque, 2012; Amengual & Mendez, 2012), but literature on pilot studies about the implementation of such proposed test remains limited (Martin-Monje, 2012) and more research should be done on the implementation of such tasks in real environments.

In this paper, we report on an experience using the proposed tasks in four different provinces in Spain with the aim of illustrating the validity of such proposal both in terms of students' results and in terms of the strengths and the constraints found in those settings. The creation and evaluation of the test is part of a wider project, the OPENPAU project (Spanish Ministry of Education, 2011-2014, FFI2011-22442), whose aim is to put forward a solid proposal to facilitate the implementation of the oral part of the exam. The analysis done in this paper is based on a wide range of factors such as testing procedures, teachers' attitudes, and test organization and delivery (i.e. pen & paper or computer based).

First, a rationale for the tasks chosen and the results of previous experiences trialling speaking tasks will be provided. Second, the organization and the methodology for this research will be explained. And finally, the key strengths and difficulties found in the four settings will be addressed and some conclusions will be drawn.

II. Literature Background

Several authors (Bueno-Alastuey & Luque, 2012; Amengual & Mendez, 2012) have proposed the characteristics which tasks should display to test students´ productive oral performance “in appropriate, contextualized, communicative language use” (Bachman, 1990, p. 84) in the Spanish context. These studies have proposed the inclusion of two tasks, an individual monologue task and a pair or group interaction task, so that both individual production and interaction as facets of oral communicative competence are tested, and thus the authenticity, reliability, and construct and content validity of the test are improved. The inclusion of prompts to “help candidates contextualize and elicit the required responses in the target language” (Amengual & Méndez, 2012, p. 117) has also been recommended.

The inclusion of both a monologue and a group based task was also based on the fact that even though group speaking tests have become prevalent (Cheng, Rogers, & Hu, 2004), some recent research (Bahrani, 2011; Martin-Monje, 2012) has illustrated the possibility of getting positive results without any synchronous interaction in computer-based pilot studies. Furthermore, pair and group tests may not always benefit students’ performance (Saito, & Miriam, 2004), as aspects like personality (Tsou, 2005), anxiety (Mohammadi, Biria, Koosha, & Shahsavari, 2013) competitiveness, discourse co-construction (Zhang, 2008; Sabet, Tahriri & Pasand, 2013), motivation, learning styles (Tuan, 2011), scales (Hudson, 2005), sex (Azkarai & Mayo, 2012), channel, tester, raters and many others have a powerful effect on the final assessment. Besides, interaction may not be the only way to trigger the candidate’s performance and, consequently, both types of tasks seem to be necessary.

Analysis of teachers´surveys regarding the kind of tasks students at this level perform orally in their classes also pointed in the same direction (Garc ía Laborda & Fernandez Alvarez, 2012) and, thus including an individual task to evaluate oral communicative competence in monologue, and a pair task to test it in interaction increases the face-validity of the test and its practicality.

Although these two tasks have been proposed and they are included in some high-stakes international proficiency tests such as the Cambridge tests, there is no research which has analyzed students´results, the organization and implementation of such test, and its feasibility in Spain.

As some research has pointed out that tasks and cultural variables may affect performance (Fulcher & Marquez, 2003), it is not enough to propose the construct and type of exercises of the test but it is also necessary to try it in experimental conditions to make sure that this test is appropriate for students in different parts of a particular setting, in this case in different regions of Spain (Weir, 2005).

III. Methodology

A. *The Project*

The OPENPAU project (see Fig.1) is currently being carried out by two groups: a computer centred sub-group (02), and a large sub-group (01) of a number of subjects grouped in much smaller groupings that include content researchers, specialists in measurement and research, the test coordinators and guest researchers. Both groups work on different aspects of the test. On the one hand, sub-group sub-group 01 works on research methodology, the construct of the test, the tasks and their analysis based on pragmatics, discourse analysis etc., their application to the students, and washback. On the other hand, subgroup 02 focuses on the technological aspects of the project, the testing platform design and its implementation and the process control.

This paper is part of the work of sub-group 01 and it is an experimental trialling of a first version of the test prior to being included in the testing platform.

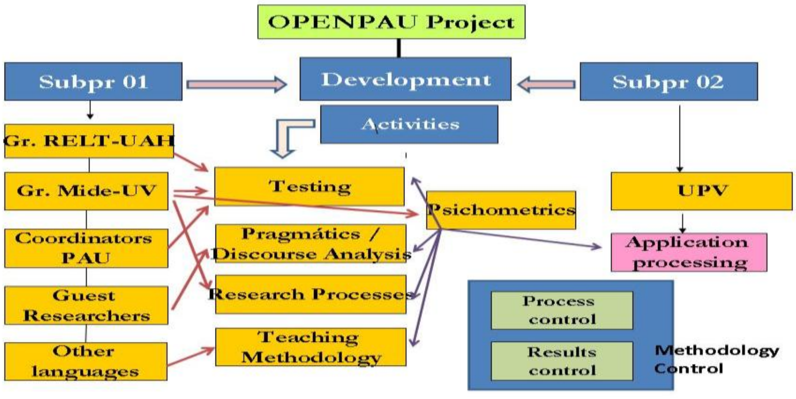

Figure 1. Coordination of the OPEPAU project

B. *Context*

Although the Baccalaureate General Test is expected to be very similar in all Spanish regions, the current teaching and learning processes taking place in those locations’ high schools are varied, and those different realities should be considered for the piloting phase to study whether different teaching contexts and realities affect the results students obtain in the test.

Although lately classroom time devoted to speaking tasks seems to have increased in all regions, probably due to the fact that the oral component of the exam was going to be included in the new Baccalaureate General Test in the year 2013, the amount of time allotted to speaking tasks in high schools differs depending on the region. For instance, the region of Madrid (the capital city) may have up to seven hours a week of English language instruction and four of those hours are devoted to speaking, while in other regions such as Navarre students receive four hours of instruction per week with no specific weekly time assigned to speaking. However, it is very difficult to establish how much time is devoted to speaking and listening in each of the regions, and even in two different schools in the same region as generally there is no record of the time devoted to any of the skills weekly. Consequently, no specific general differences can be assumed between specific regions.

To overrule the possible limiting effect of studying only a specific context, four provinces were selected for our pilot study to represent as much variety as possible. Therefore, we selected a small community in the north of Spain, Navarre, a big community in the south of Spain, Andalusia, and two provinces in the middle of Spain, Castile - Leon and Castile - La Mancha, which includes the capital of Spain.

C. *Participants*

The participants were 168 students either studying the last year of secondary high school (so a few months before having to take the University Entrance exam) or the first year of a university degree (having just done the University Entrance examination the previous year). 28 students were studying in Navarre, 38 in Andalusia, 62 in Castile-Leon and 40 in Castile-La Mancha.

D. *Instrument*

As it has been mentioned, the pilot speaking test considered the usefulness criteria established by Bachman and Palmer (1996), and included the type of tasks recommended by previous studies to fulfill the criteria of validity, reliability, authenticity, wash-back effect and practicality (Bueno-Alastuey & Luque, 2012).

It included 3 parts. The first part was an introductory conversation between examiner and test takers. The test takers were asked to provide personal information about themselves (likes, dislikes, hobbies etc.), their experiences learning English and some background information about previous exams taken and abroad experiences. It included questions such as (1):

(1) *What did you do this morning/yesterday?*
What do you do in your free time?

This part of the test was used to try to break the ice and to give the students the opportunity to relax.

The second part consisted of a monologue based on a description of two pictures (different for each test taker). It included questions such as (2) :

(2) *Describe what you see in the picture.*
Would you like to be in this picture doing this activity?
What are the advantages and disadvantages of each activity?

Finally, the third task was a discussion in pairs about their personal preferences between two given situations. It included questions such as (3) :

(3) *Describe your ideal place to live / Which do you prefer? The city or the country? Why?*
Describe where you live now.

E. *Data Collection and Analysis*

The students did the interviews in a room with different examiners. All the interactions were recorded. The monologues and interactions were judged according to the levels illustrated in the Common European Framework of Reference for Languages (A1, A2, B1, B2, C1 and C2), where level A1 corresponds to beginner, A2 to pre-intermediate, B1 to Intermediate, B2 to Upper-Intermediate, C1 to Advance and C2 to Proficient. All levels are described according to competence in the four skills and sociolinguistic competence, and those competences should be tested communicatively and based on performance (Council of Europe, 2001).

The analysis was based on students' results in the test and on a categorization of the main strengths and weaknesses of the tasks proposed together with students´ difficulties while doing the tests. The categorization was based on examiners reports about the procedure and the tests.

IV. Results and Discussion

A. *Navarre*

The conversations were scored based on four factors: fluency, accuracy, interaction and coherence. From the 28 students who took the test, 25 were considered to have a level B1 or above, 2 a level A2 and 1 a level A1. Agreement between raters was high and the reliability of the test and the scores was further supported by the fact that the teachers of the students confirmed their students had performed as expected.

The main strengths of the speaking test were three. First, the fact the students carried it out in their own schools with known peers so they were familiar with the context and with their interlocutor. Secondly, as their regular teacher was also present, the anxiety this kind of tests provokes was minimised by the familiar environment and interlocutors. Thirdly, the fact that the tasks included a monologue, in which students showed their competency level, as well as a dialogue, in which they could help and be helped by prompts by their peers, could also offer a more realistic picture of their communicative oral competence.

The main difficulties of this kind of test were the pressure of being recorded and the fact that it had to be carried out outside regular classroom hours and this meant that only volunteers were tested. This might mean that students with lower speaking proficiency levels did not volunteer and so the group might not be representative of the real amount of students with at least a B1 level of English.

Students performed well in the two kinds of tasks proposed (monologue and pair interaction) and they seemed to be familiar with both types of tasks.

Consequently, we can conclude that monologues, descriptions of pictures and discussions with visual support are the most common type of tasks practiced in classrooms. This finding coincides with previous research stating teachers´ preferences for this kind of tasks for oral development and testing (Martinez, Sevilla & Gimeno, 2009). Regarding level, students performed at a B1 level of spoken competence as described in the CEFR (Council of Europe, 2001).

B. Andalusia

Students results in the pilot test ranged from an A2 level to a B1 or beyond, with some exceptional cases of an apparent A1, perhaps because of attrition or fossilization phenomena which would be solved with short periods of intensive instruction (Muñoz, 2012).

The strengths of the test were the design of the tasks as students scored better in direct questions and after speaking for several minutes. Consequently, the inclusion of several tasks and one of them with direct questions was considered valid. Another strength was the fact that the tasks included visual or verbal support, so students´anxiety level decreased, and communication became more effective. In fact, students tended to do better at fluency than at accuracy, but their communication skills lowered when interacting with someone else. Explicit teaching of common vocabulary and communication strategies together with frequent practice at interaction would likely improve competence levels.

The difficulties students experienced in the test included feelings of embarrassment because of self-perceived lack of communicative skills due to lack of practice in their English classes, where they seemed to have had more practice in answering than in interacting. Secondly, they also showed lack of vocabulary for everyday communication. Thirdly, a frequent lack in the use of those communication strategies that deal with simplification or elaboration phenomena was observed, supporting previous research findings (Manchón, 2008). Furthermore, formulaic language, self-monitoring or use of adequate fillers (Bygate, 2001), which would promote students´oral skills, was not noticed either.

An explaining factor for the lack of communicative skills perceived in the piloting test can be the teaching method that seemed to have been used in their high schools English lessons. With a very high average number of students per classroom, not sufficient opportunities for oral practice or interaction seemed to have been provided. Besides, as the traditional Baccalaureate General Test only tests reading, grammar and vocabulary, the washback effect has been a basically teacher-directed methodology focussing on writing and reading skills in their learning context.

C. Castile & Leon

62 students took the test in Castile-Leon. Overall, the speaking test designed proved to be suitable –with slight adjustments- to determine whether a high school senior student had a minimum speaking skill of Intermediate B1 (threshold level) before entering university. The results of the students ranged from A1 (7 students), A2 (23), B1 (28) and B2 (5), thus the majority of students were in the A2 - B1 range, which confirms previous findings regarding the level of first year university students in Spain (Bueno-Alastuey & Ballarín, 2011)

Regarding the students' performance, the task where they did more poorly was the third (i.e. the dialogue in pairs). In many cases, students found it difficult to make up a conversation and talked individually in turns instead. This fact resulted in a low grading of the "interaction" aspect of speaking.

The strengths of the test can be classified in three major groups: materials used, the procedure itself and the instructor's assessment. Regarding the materials used, the fact that the pictures were in colour was considered a benefit because all details were easier to be distinguished and the students were given the opportunity to describe the photos in a more complete and accurate way, provided they had enough vocabulary to do it. The fact that the examiner had a bank of questions for the first and second part of the test was also considered positive. Lastly, the introduction of two types of tasks (a monologue and pair interaction) made the testing of communicative competence more complete.

Regarding the procedure itself, it was an advantage that the students had to do the test in pairs with familiar partners. This fact is important, especially among Spanish students, considering the unconscious barriers which prevent them from speaking in a foreign language. Research has also confirmed that acquaintance with speaking partners has a positive effect on speaking performance (O'Sullivan, 2002)

Regarding the assessment, having an assessment rubric with four aspects - fluency, accuracy, interaction and coherence - allowed the examiners to score speaking proficiency on the same terms.

The difficulties found while carrying out the oral tests were also related to the materials used and the assessment. In the first case, the examiner's bank of questions for the first and second parts should be a bit more ample to guarantee the variety of topics which students can talk about in order to prevent repetition. As for the third task, a sketch or simple drawing could be more suitable to pose the facts and different options the students should discuss before making a decision. The description of the pictures could be optimized if students were asked to describe the people as well as the setting or the actions.

As far as the assessment is concerned, a more detailed rubric introducing aspects such as linguistic repertoire, vocabulary, or pronunciation would help the examining teachers to calibrate the student's real speaking skill in a more accurate way.

D. Madrid & Castile La Mancha

In Madrid, the most striking finding was the great difference between those who had had 4 hours of instruction and those who had had extensions, which means 7 hours of instruction. While some of the students evidenced an adequate B2 competence, others could hardly achieve an A2. Most students underscored and were ranked in between A1 and A2 despite having had a large number of years of English language instruction.

The strengths of the test were the kind of tasks. They were adequate for the purpose of finding about students´ general competence and included monologues and dialogues, thus, measuring interaction. Students showed more proficiency on communicative/interactional tasks than on descriptive or teacher-student tasks. Consequently, the design of the Baccalaureate General Test should aim at interaction.

The weakness noticed was the limited use of Cognitive Academic Language Performance skills. Since the use of English by university students should be related to both their personal relationships and academic work, the almost complete absence of part of this knowledge was considered a major issue. In reference to the tasks, tasks oriented towards professional and technical communication should be incorporated.

V. CONCLUSIONS

After analysing the findings of a pilot test for the proposed future speaking section of the Baccalaureate General Test in four geographically dispersed settings in Spain, the following conclusions about the strengths and shortcomings of these proposed tasks could be drawn and the following modifications put forward to improve the test design and implementation.

The construct of the test can be considered valid regarding the type of tasks and the materials used according to recommendations from previous research in the field (Bueno-Alastuey & Luque, 2012). Regarding the procedure, the pairing of students familiar to each other and the presence of the teacher were reported as favourable for students to show their competence. The fact that students had to show both their competence in a monologue and in dialogues, thus their interactional competence, was also judged as positive and reliable. Finally, using a rubric to evaluate oral performance also increased the consistency of examiners ´scores.

The level of students, ranging between A2 and B1 in most regions, seemed to be enough to pass this kind of oral test, although some examiners pointed out negative feelings of embarrassment and inadequacy in some students and lack of communicative competence strategies to deal with communication more successfully. These negative factors could be overcome by an increase in the time devoted to the speaking skill in class and by explicit instruction of communicative competence strategies in high school classrooms. In some regions students performed better at interaction, and in others in monologues, so both kinds of tasks should be included to get a more reliable and valid score of speaking competence.

The main reported shortcomings of the test were related to the limited quantity of questions, according to some of the examiners, in the first part, and to the modest quantity of visual support. Consequently, some of the raters suggested more questions in the first part, more visual support in the second part and some guide and visual support, for example graphs, in the discussion part of the test. A more detailed rubric including more aspects such as linguistic repertoire, vocabulary and pronunciation were also demanded together with a broader band of scores.

As a conclusion we must point out that the aforementioned recommendations should be included in further tests and then piloted to measure satisfaction. Further research should also be initiated on the implementation of a speaking test using technological tools so as to improve the economy in terms of time of the speaking section of the future Baccalaureate General Test.

ACKNOWLEDGEMENT

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 ERDF funds under the 2008-2011 plans.

REFERENCES

[1] Amengual, M. & Méndez, C. (2012). Implementing the Oral English Task in the Spanish University Admission Examination: and International Perspective of the Language. *Revista de Educación*, 357, 105-128.

[2] Azkarai, A., & Mayo, M. d. P. G. (2012). Does gender influence task performance in EFL? interactive tasks and language related episodes. *Utrecht Studies in Language and Communication, 24*, 249-278.

[3] Bachman, L.F., & Palmer, A.S. (1996). Language Testing in Practice: Designing and Developing Useful Language Tests. Oxford: OUP.

[4] Bahrani, T. (2011). Speaking fluency: Technology in EFL context or social interaction in ESL context? *Studies in Literature and Language, 2*(2), 162-168.

[5] Bueno-Alastuey, M. C. & Luque, G. (2012). Competencias en lengua extranjera exigibles en la Prueba de Acceso a la Universidad: Una propuesta para la evaluación de los aspectos orales. *Revista de Educación*, 357, 81-104.

[6] Bueno-Alastuey, M.C. & Ballarín, A. (2011). Capacitación lingüística de ingreso en la UPNA. Un estudio comparativo de varias titulaciones de grado. *Huarte de San Juan. Filología y Didáctica de la Lengua*, 11, 73 -91.

[7] Bygate, M. (2001). “Speaking”, en R. Carter & D. Nunan (eds.) *The Cambridge Guide to Teaching English to Speakers of Other Languages*. (pp. 14-21). Cambridge: Cambridge University Press.

[8] Cheng, L., Rogers, T., & Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing, 21*(3), 360-389.

[9] Council of Europe. (2001). Common European framework of reference for language learning and teaching. Cambridge, UK: Cambridge University Press.

[10] Ekbatani, G. (2011). Measurement and evaluation in post-secondary ESL. New York: Routledge.

[11] Fulcher, G. & Marquez, R. (2003). Task difficulty in speaking tests. *Language Testing,* 20 (3), 321 – 344.

[12] Fulcher, G. (2010). Practical language testing. London: Hodder Education.

[13] García Laborda, J. & Fernández Álvarez, M. (2012). Actitudes de los profesores de Bachillerato de Alcalá y Navarra ante la preparación y efecto de la PAU. *Revista de Educación*, 357, 29-54

[14] Hudson, T. (2005). Trends in assessment scales and criterion-referenced language assessment. *Annual Review of Applied Linguistics, 25*, 205-227.

[15] Manchón, R.M. (2008). Taking strategies to the foreign language classroom: where are we now in theory and research? *IRAL* 46, 221 – 243.

[16] Martín Monje, E. (2012). La nueva prueba oral en el examen de inglés de la Prueba de Acceso a la Universidad. Una propuesta metodológica. *Revista de Educación, 357*, 143-162.

[17] Martinez, A, Sevilla, A. y Gimeno, A. (2009). *Resultados encuesta profesores de Bachillerato nueva prueba de lengua extranjera PAU.* Internal report from the University of Valencia available at http://www.upv.es/ingles/documentos/informe.pdf.

[18] Mohammadi, E. G., Biria, R., Koosha, M., & Shahsavari, A. (2013). The relationship between foreign language anxiety and language learning strategies among university students. *Theory and Practice in Language Studies, 3*(4), 637-646.

[19] Muñoz, C. (ed.) (2012). Intensive Exposure Experiences in Second Language Learning. Bristol: Multilingual Matters.

[20] O’Sullivan, B. (2002). Learner acquaintanceship and oral proficiency pair-task performance. *Language Testing,* 19(3): 277–95.

[21] Sabet, M. K., Tahriri, A., & Pasand, P. G. (2013). The impact of peer scaffolding through process approach on EFL learners' academic writing fluency. *Theory and Practice in Language Studies, 3*(10), 1893-1901.

[22] Saito, H., & Miriam, E. E. (2004). Seeing English language teaching and learning through the eyes of Japanese EFL and ESL students. *Foreign Language Annals, 37*(1), 111-124.

[23] Tsou, W. (2005). Improving speaking skills through instruction in oral classroom participation. *Foreign Language Annals, 38* (1), 46-55.

[24] Tuan, L. T. (2011). EFL learners' learning styles and their attributes. *Mediterranean Journal of Social Sciences, 2*(2), 299-320.

[25] Weir, C.J. (2005). Language testing and validation. Basingstoke: Palgrave Macmillan.

[26] Zhang, L. J. (2008). Constructivist pedagogy in strategic reading instruction: Exploring pathways to learner development in the English as a second language (ESL) classroom. *Instructional Science*, 36(2), 89-116.

Mª Camino Bueno-Alastuey, PhD, is a lecturer at the Public University of Navarre, Spain, where she teaches English Teaching Methodology courses related to ICT for learning and teaching foreign languages to Education Degree and Master students and English for Specific Purposes. Her research focuses on CALL, blended learning design and implementation, and the use of technological tools for testing oral competence. She has published nationally and internationally on these topics and is a regular reviewer for national and international journals.

Jesús García Laborda is an associate professor at the Universidad de Alcala (Madrid, Spain). Dr Garcia Laborda has a PhD in English Philology and an EdD in Language Education. His current research covers many areas of computer implementations for language learning and testing along with ESP and teacher training, especially applications of low stakes online testing, especially focused to the Spanish University Entrance Examination, mobile learning for commercial purposes to help to introduce a computer based test in high stakes exams such as DELE (Spanish) or the Spanish University Entrance Examination (English, French, German), and the implications of implementing such test in teacher training along with more traditional approaches to teacher education, and trainees’ development of both cognitive and computer skills.

Ana Isabel Muñoz Alcón is an English teacher in the Sciences and Arts Faculty of the Catholic University of Avilá, where she leads a team of interdisciplinar researchers focusing on Formative Evaluation. Previously, she was an English teacher for 14 years working in Secondary Education, where she researched various innovative evaluation projects in the field of Foreign Languages.

Gloria Luque Agullo is an associate professor at the University of Jaén. Her research deals with the field of Applied Linguistics and Teaching Methodologies and materials. She has also wide experience in language testing, participating in the evaluation of University Entrance Exams for 11 years. For the last 8 years, she has also been a coordinator of this exam in the region of Andalucía. Her present/continuing research deals with how oral language is developed in textbooks.