BRITISH COUNCIL

ENGLISH LANGUAGE
ASSESSMENT RESEARCH GROUP

STUDENT PERCEPTIONS OF THE CEFR LEVELS AND THE IMPACT OF GUIDED PRACTICE ON APTIS ORAL TEST PERFORMANCE

AR-G/2017/5

Jesús García Laborda, Marián Amengual Pizarro, Mary Frances Litzler, Soraya García Esteban, Nuria Otero de Juan

ABSTRACT

This study compares students' scores on an Aptis Speaking Test to their results during an informal oral interview and a mock speaking test. First, it examines differences in the testing of oral skills between in-person paired interviews and a computer-delivered test, both following the British Council Aptis test, the former rated by university professors and the later rated by Aptis. Second, it analyses whether there are significant differences in performance due to repetition of the test and guided practice. In addition, it looks at the participants' different skill scores from the test to determine whether there are any consistent variations within the group. Finally, the study examines the students' impressions of their level on the CEFR. A total of 86 students took part in this exploratory study that has a quasi-experimental design. The data for the study include initial speaking interviews conducted by the research team, Aptis test scores, and qualitative along with quantitative questionnaires.

The study found that guided practice and repetition of the test lead to slightly higher scores, and that there were mixed results compared to the two measures.

Authors

Jesús García Laborda is an Associate Professor at the University of Alcalá. He has a Doctorate in English and another in Language Education. He has focused his research on computer-based language testing, computer-assisted language learning and teaching, English for Specific Purposes and teacher education. His publications in testing include articles in *Computers & Education, Educational Technology & Society, Language Learning & Technology, Universal Journal of Computer Science* and more. He has done extensive research on the Spanish University Entrance Examination and co-operated with a number of institutions, particularly the Spanish Ministry of Education.

Marián Amengual Pizarro is an Associate Professor at the University of the Balearic Islands and Vice-Dean of the Philosophy and Letters Faculty. She has a Doctorate in English. Her research includes language testing and teacher education. She is one of the editors of *Revista Electrónica de Lingüística Aplicada* (RAEL). Her latest publications have been included in *Vigo International Journal of Applied Linguistics* (VIAL), *English Studies* and *The European Journal of Teacher Education*.

Mary Frances Litzler is an Assistant Professor at the University of Alcalá. She teaches English language, linguistics and translation to students in the English Studies, Modern Languages and Translation, and Education programs of study. Her major research and publications deal with diachronic linguistics, language teaching and learning, learner autonomy and computer-based language testing. She has taught English language in Spain, Japan and the United States.

Soraya Garcia Esteban holds a Ph.D. in Modern Languages, Literature and Translation from the University of Alcalá and an M.A. from the University of the West of England, Bristol. She is an undergraduate and postgraduate lecturer at Cardenal Cisneros and the University of Alcalá (Spain). She is involved in various European and international research projects and has published in the area of lexicography and microteaching practice in infant and primary education. Her research interests are CLIL, ESP and promoting language learning with technology.

Nuria Otero de Juan has an M.A. in Spanish Education and extensive experience as a researcher in language testing. Her research interests are language learning with technology and discourse interaction.

CONTENTS

Figure

Tables

1. INTRODUCTION

English is a fundamental language for commerce and education, as well as other spheres of activity in Europe and around the world. Following European Union directives for European citizens to master two foreign languages by the year 2020, Spain has placed increased importance on the teaching of English by setting up bilingual education programs in many schools. For example, 40% of all the schools in the Community of Madrid are now bilingual.

University students in Spain are now also normally required to certify their level of English in order to graduate. At the University of Alcalá, for example, students must demonstrate a B1 level of the Common European Framework of Reference (CEFR) for the majority of the degree programs, but students studying specific fields related to English have other requirements, as in the case of Education with a specialisation in English (B2) and Modern Languages and Translation (C1).

An efficient, fast measurement process is needed to meet these demands each year but the use of in-person evaluators can be very slow and cumbersome, and many teachers who serve as evaluators do not receive any compensation. For this reason, an online test can be highly useful, especially if it can be offered at an affordable price. The Aptis testing system is a possible way to meet this need. According to O'Sullivan (2012), this testing system is flexible and accessible. In addition, the scores follow the CEFR and the test can be adapted to the needs of organisations. In Spain, for instance, it is used for certification of bilingual education teachers.

However, in recent years, a number of Aptis candidates at the University of Alcalá have obtained CEFR levels B2 or C1 in all the skills except for Speaking, which has been rated as low as A2 or B1. In addition, in 2013–14 some professors at the university became concerned about discrepancies between Aptis Test scores and their own observations of student levels in class throughout the academic year. At the same time, students have commented on differences between delivery modes for oral exams. In fact, differences in level have been observed between face-to-face oral exams and computerised exams, such as Aptis. This has especially been the case in the English Studies program at the University of Alcalá, where students are required to certify a level of C1.

Justification of differences and inferences in scores across skills in high-stakes exams has been an issue for more than 50 years. In the case of the Department of Modern Philology at the University of Alcalá, the observed discrepancies could have been due to a number of factors, including the training, preparation and experience in testing of the professors, as well as the students' possible lack of familiarity with computers. The students in the English Studies program tend to avoid using computers at the university, a situation that is reinforced by the fact that fewer than 10% of the courses offered in the program use computer tools on a regular basis. For all of these reasons, it was imperative to investigate whether these two factors were the cause for the discrepancies observed.

This study follows a quasi-experimental design to determine the effectiveness of the Aptis Test in relation with different contextual variables. In other words, it attempts to shed light on the question of why students score lower on the Aptis Speaking Test compared to the other sections. In particular, it aims to shed light on answers to the following questions:

1. Do university professors who rate students assign the same level of oral proficiency as the Aptis raters?

2. Does knowledge of the different tasks in the Aptis Speaking Test and the opportunity to have time to prepare for the Speaking test affect student results on this part of the test?

3. What do the students perceive their speaking level to be? How do their perceived levels compare to their actual Aptis Speaking Test scores?

4. Do the students obtain mixed level scores for the different sections of the Aptis Test as has been observed in recent years at University of Alcalá? If so, do any skills stand out as being consistently lower or higher among the participants?

2. THEORETICAL FRAMEWORK

In the last century, it was common to believe that students showed similar levels of competence in the four traditional skills, but Oller (1979) began to change this controversial idea. Since Weir's (2005) description of the framework of language testing, a number of issues have been addressed in context – delivery being a significant one in recent times. Today, variations in skill level are thought to be due to a variety of factors including: 1) different perceptions on the part of test-takers as to the performance characteristics needed to obtain certain scores (He & Shi, 2008); 2) a mismatch between their expectations and the actual test itself (Xie & Andrews, 2013; Simpson, 2006) and, finally; 3) human-computer interaction during test delivery and the anxiety that arises when limited training possibilities are available, as in the case of the Aptis Test (Van Moere et al, 2009). These three aspects require serious examination.

3. METHOD

3.1 Overview

The study took place in two phases. In Phase 1 the students were interviewed by the researchers to determine their levels of spoken English. In Phase 2 they were separated into a control and an experimental group. The experimental group received information about the Aptis Speaking Test and materials to practice for a speaking test. After two to three weeks, both groups of student took a first official Aptis Test. A selection of the students from the experimental group was then asked to study more for a second official Aptis Test, which they took three weeks later. During the different phases of the study the students filled out three questionnaires. The first phase took place between late November 2015 and February 2016. The second phase lasted from March 2016 through late April 2016.
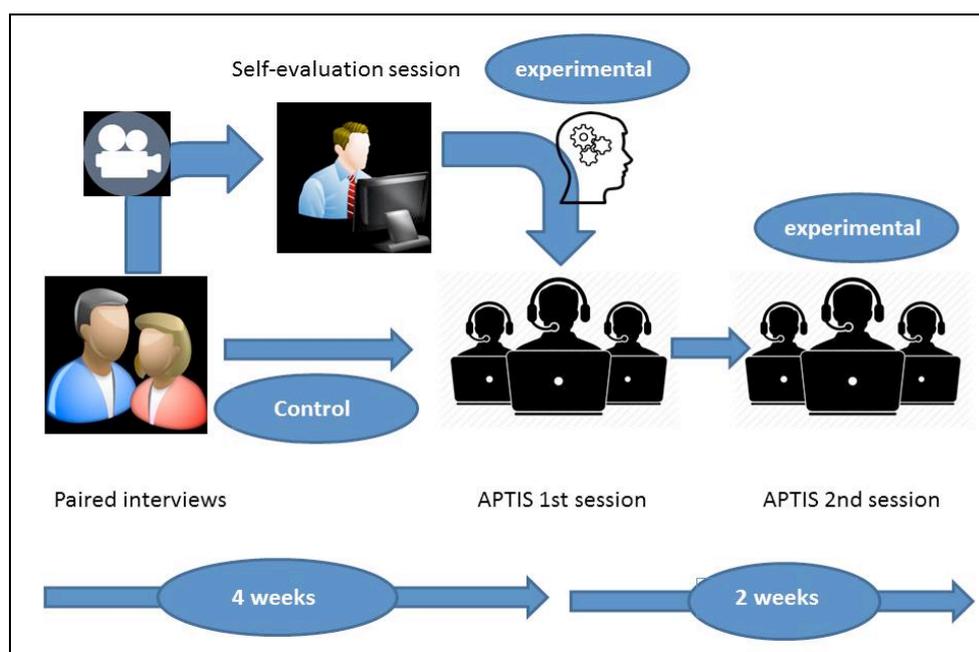

Figure 1: Overall project

3.2   Subjects

A total of 86 students took part in this study including those who were unable to complete all of the phases. Of the students, 56 were from the Faculty of Education. These students did not have experience with the Aptis Test prior to this research project. Before completion of their degrees in Education, they must certify a B1 level on the CEFR, but if they would like to opt into a special course with additional classes in English, they must certify a B2 level by the end of the second of the four years of study. This means that this group of students was expected to have an English level in the range of A2–B2 at the time of the experiment.

An additional 30 students were from the Faculty of Arts and Humanities, where they were enrolled in English Studies or Modern Languages and Translation. These two programs require certification of a B2 level upon completion of the first year of study and a C1 level at the end of the four years, meaning that these students were expected to have a B2–C1 level at the start to this research project. Some of these students had experience with the Aptis Test because they used it to certify their level in the first year at university. All of the students participating in the research project signed up voluntarily as they knew that they would be able to take an Aptis Test that could be used for their respective certification requirements.

Variable	Responses	Experimental group		Control group	
		Frequency of answer	Percentage %	Frequency of answer	Percentage %
Number	**Total**	**56**	**100%**	**30**	**100%**
Age in years	18–20 years	22	39.9	10	33.3
	21–22 years	19	33.9	7	23.3
	23–25 years	6	10.7	7	23.3
	Over 25 years	5	8.9	3	10
	Not answered	4	7.1	3	10
Gender	Male	12	21.4	4	13.3
	Female	44	78.6	26	86.7
Degree program	Early Childhood Education	1	1.8	6	20
	Primary Education	20	35.7	21	70
	Modern Languages & Translation	4	7.8	0	0
	English Studies	26	46.4	0	0
	Not answered	5	8.9	3	10.0
Year of study	2nd	19	33.9		
	3rd	23	41.1	26	86.7
	4th	7	12.5	1	3.3
	Not answered	7	12.5	3	10
Study of English in years prior to attending university	5–7 years	2	3.6	6	20.0
	7–12 years	23	41.1	7	23.3
	More than 12 years	27	48.2	13	43.3
	Not answered	4	7.1	4	13.3

Table 1: Descriptive variables for the two groups

Background information on the two groups, as obtained from a first questionnaire (Appendix 2) can be seen in Table 1. Both groups were largely made up of female students between the ages of 18–22 who had studied at least seven years of English before attending university and in many cases more than 12 years. Many of them were in their second or third year of university.

3.3.   Procedure

3.3.1   Initial interviews

Initial interviews of the students were conducted in pairs in November 2015 and in early February 2016 in order to determine the students' levels. Three of the research team members worked separately to ask the students a series of questions similar to questions that can be found on the Aptis Speaking tasks and IELTS Speaking section. This first phase was conducted with no opportunity for the students to prepare, and the oral interviews were carried out by pairing the students randomly. The tasks included warm-up questions on a personal topic, a description of a photograph, a comparison of two photographs, and a discussion of subjects that ranged from concrete ideas to more abstract topics. Half of the tasks were individual, as in the case of the Aptis Test, but the other half were completed as a conversation in pairs to determine the students' ability to interact with each other. A list of sample prompts used can be found in Appendix 1. These initial interviews were recorded on video so that the students' non-verbal dynamics can be studied at a future date, even though the Aptis Test is only audio-recorded. Once all of the recordings had been made, the three members of the research team proceeded to score them according to the CEFR levels. Upon completion of the interviews, each student filled out a first questionnaire (Appendix 2).

3.3.2.   Informational sessions

In mid-March, the students in the experimental group each attended one class session of approximately one hour and fifteen minutes during which they evaluated their own level of English as observed in the initial interview, answered a second questionnaire (Appendix 3) and received information and tips about the Aptis Speaking tasks. During the first part of the sessions, which took place in the computer laboratory of the faculty, the students were provided with links to view videos for the B1–B2 and C1 CEFR levels. The videos were taken from the YouTube channel, "CambridgeeducationTV" and they show people doing the oral test of the different Cambridge exams. The students viewed parts of these videos during the informational session, and were then given a USB drive with their interview video so that they could evaluate their own performance in terms of different aspects of English and the CEFR levels and bearing in mind the model videos. They wrote their impressions as part of the questionnaire, which they filled out and submitted during the informational session.

The last part of the class session with the experimental group was an explanation of the format of the Speaking section of the Aptis Test, along with tips for success on the exam. The information was based on the "Aptis Candidate Guide" found on the Aptis website. The students were told where they could find the guide later if they wanted to consult it after the session. At the end of the class period, the students were provided with test preparation materials for the IELTS Speaking section so that they could practice prior to the official exam. Materials for this other exam were selected because no materials specifically designed for Aptis Test preparation were found. It was pointed out to the students that the IELTS Speaking Test has a different format but that there were some interesting tips in terms of phrases and language for use during an official exam and some ideas on testing strategies, along with possible test questions that could serve to give them ideas for practicing speaking. The students were also asked not to discuss or share the material with others as they were participating in an experiment.

In mid- March, the control group students also attended a class session in the computer laboratory of the Faculty of Education. The two sessions held to accommodate this larger group were somewhat similar to the experimental group class session in that the students viewed the same model speakers for the different CEFR levels and they answered the same questionnaire, excluding the self-assessment questions. In the case of the control group, however, the students viewed clips of several minutes from the videos as a group, as opposed to individually at their computers. They were also not provided with any information on the exam structure itself. However, it should be pointed out that these students were taught exam strategies as part of their English classes at the university. These included awareness and self-evaluation of their range and accuracy of structures, fluency, and interaction, etc. To ensure that the information explained during the group sessions was consistent, the same researcher addressed the different groups using the same PowerPoint presentation. These sessions were also recorded so that the research team could consult them as needed.

3.4.   Testing and experimentation

The official Aptis Test was delivered to all of the students on 1 April and 8 April 2016 in a morning or afternoon session (two to three weeks after the information sessions depending on the group). All of the students, including those who had received practice materials as well as those who had not, took the tests during these two weeks. The Aptis Tests were delivered under official conditions in a computer room used by the British Council to deliver Aptis Tests for the university on a regular basis. Four different sessions (morning and afternoon) were held for the test as the room has only 24 computers.

3.4.1. Mock Speaking Test

In the hour before the official exam, the students were given a mock speaking test developed by the research team, which followed the types of questions explained in the "Aptis Candidate Guide". A photocopy with questions similar to those found in the guide (Appendix 4) was provided to the students and they were allowed 15 minutes to read the questions and to prepare as they wished, knowing that they would have to record their answers afterwards. After the 15 minute preparation period, one of the researchers (a native speaker of English) read the questions out loud to the group as they read along silently and she timed the students following the times permitted for answers during an Aptis Test, while the students recorded themselves using the language laboratory computers. The computers are equipped with headphones and microphones and they are located in cubicles to prevent environmental sound from interfering with the recording quality. Upon completion of the simulation, the recordings were collected by the researchers using the laboratory software; they were later scored by trained Aptis raters.

3.4.2. Aptis Test 1

After the mock exam, the students were given a short break before continuing to the laboratory in the next room, where they took the complete Aptis Test. As mentioned above, the test was delivered under official conditions by a British Council examiner who often runs Aptis Tests in the same laboratory. The students answered a third and final questionnaire (Appendix 5) online in the week following the official testing session.

3.4.3. Aptis Test 2

A selection of students who had obtained lower than expected scores or who had slightly more uneven scores for the different skills on the first Aptis Test was made upon receipt of the scores from the British Council. These students were asked to study the test preparation materials for the following two weeks and to return on 29 April for another complete Aptis Test. Thirty-one (31) students took the exam this second time. More students were not included as they were unable to continue in the project due to their university commitments.

3.4.4. Questionnaires

The participants filled out three questionnaires during the research project.

3.4.4.1. Initial questionnaire

Upon completion of the initial interviews, the participants filled out a questionnaire on their personal details (age, sex, experience with English study and exams, etc.), their impressions of their English level during the initial interview and in general, and their ideas about the Aptis Test and the levels of the CEFR. This last group of questions was designed to see what the students knew or did not know about the CEFR and the Aptis Test. An English translation of the questionnaire, which was issued in the students' native language, Spanish, can be found in Appendix 2.

3.4.4.2. Second questionnaire

A second questionnaire was answered during the informational class sessions. The first section had questions to prompt the students to evaluate their performance during the initial interviews in comparison with the models provided on the video links described in Section 3.3.3 above. These were quantitative questions regarding their impressions of different aspects of their English ability, such as pronunciation, structures and ability to communicate their ideas. In addition, the questionnaire asked open-ended questions about their past experience with instruction regarding the CEFR levels, their test-taking strategies, ideas on how to improve, and affective aspects of test-taking. An English translation of this second questionnaire can be found in Appendix 3.

3.4.4.3. Final questionnaire

Upon completion of the first Aptis Test, the students were asked to complete an online questionnaire by the following week. The questions in this case asked about differences between the initial interview and the Aptis Speaking Test, the utility of the mock test prior to the actual test itself, their use of notes during the last Speaking Test task, and affective concerns in relation to computers and testing (Appendix 5 has an English version of the questions).

4.   DATA ANALYSIS

The data collected during this project are as follows:

1.   video recordings of the initial interviews with the students

2.   scores for the Aptis Test

3.   responses to the three questionnaires by each of the candidates who completed the project.

A follow-up forum and interviews were planned, but these activities have not been held to date.

The analysis involved the following aspects:

1.   *A correlation of the levels assigned to the students by professors at the University of Alcalá and the levels they obtained on the official Aptis Test.* Three members of the research team who are also professors at the University of Alcalá, with 50 years' experience in evaluation and testing between them, divided the recordings of the students randomly among themselves; each used the CEFR bands to give the students a level based on their performance during the initial interview.

    These levels were then compared to the Aptis scores (Aptis 1 below) to determine whether they were higher, lower or the same. In addition, the results were calculated for the two different faculties where the students studied to see if there were any differences in the two populations.

2.   *Analysis of the Aptis Test results for each group.* Comparisons were made between the control group and the experimental group and between the two different faculties using descriptive statistics. Particular attention was paid to the speaking results to determine if any improvement was observed between the first official Aptis Test and the second one.

3.   *A comparison between the mock speaking test and the actual Aptis Speaking Test was made using descriptive statistics to determine any differences in the scores obtained.* An overall comparison was made, as well as comparisons based on quartiles. Both the mock test and Aptis Test were rated by trained Aptis raters, as opposed to the research team members.

4.   *The levels of the students who showed improved overall Aptis Test scores during the second official testing session* were compared with their self-assessments from the second questionnaire (Appendix 3) to determine whether the students' impressions of their own performance were in line with the official test scores. The comments from the qualitative responses on the questionnaire in this regard were also examined.

5.   RESULTS

5.1   Comparison of speaking levels assigned by professors and by Aptis raters

The first research question involved a comparison of the University of Alcalá professors' scores and the Aptis Speaking Test scores. It should be recalled that the former was a spontaneous face-to-face situation while the latter was a computer-delivered one. A total of 71 students (from 78) completed both parts and could be considered for this part of the study. An overall comparison of the scores can be found in Table 2.

Aptis score- face-to-face interview score	N
Aptis score < face-to-face interview score	26
Aptis score > Face-to-face interview score	31
Aptis score = Face-to-face interview score	14
Total	**71**
Missing values	7

Table 2: Comparison of university professors' scores and Aptis Test scores

These results show that there was a difference between the two results in 63 cases and that the scores were the same in only 15 cases. In addition, the Aptis score was higher than the face-to-face score more often than it was lower. Nevertheless, contrast statistics indicated that the difference between the computer score and the face-to-face interview score was not significant for the overall group (Z=-1.260, asymptotic significance (bilateral) = 0.208). Tables 3 and 4 show the numbers of students rated at each level according to their university degree programs.

Degree – Education or English			Frequency	Percentage	Valid percentage	Accumulated percentage %
Education	**Valid values**	A1	2	4.2	4.5	4.5
		A2	5	10.4	11.4	15.9
		A2+	2	4.2	4.5	20.5
		B1	2	4.2	4.5	25.0
		B1+	7	14.6	15.9	40.9
		B2	13	27.1	29.5	70.5
		B2+	5	10.4	11.4	81.8
		C	8	16.7	18.2	100.0
		Total	44	91.7	100.0	
	Missing Values		4	8.3		
	Overall total		**48**	**100.0**		
English	**Valid values**	B2-	10	33.3	34.5	34,5
		B2	7	23.3	24.1	58,6
		B2+	6	20.0	20.7	79,3
		C	6	20.0	20.7	100,0
		Total	29	96.7	100.0	
	Missing values		1	3.3		
	Overall total		**30**	**100.0**		

Table 3: Face-to-face interview levels according to degree program

A closer look at the results (Table 5) revealed that the students from the Faculty of Education showed a different pattern from those in the Faculty of Arts and Humanities. In particular, the former group obtained a slightly higher score for the face-to-face exercise with the professors than for the computer-delivered test even though the difference was not statistically significant. The latter group, however, showed the opposite pattern as they obtained a statistically significant difference in scores in favour of the computer test.

The most outstanding finding in terms of the data comparing the university professors' ratings and the Aptis Speaking scores is that there is little consistency in the two. The result favouring the computer-delivered task at the Faculty of Arts and Humanities suggests that the students' observed avoidance of computer use at university does not seem to have affected their ability to perform during the exam. It also confirms the impression held by a group of professors in the Department of Modern Philology in 2012–13, when they voiced concerns that their students' performance in class was lower than the actual Aptis Test scores. To explain this finding, one might be tempted to think that the professors evaluated the students more harshly than the Aptis raters, but the three raters used in this case have had extensive training and experience with level testing. Nevertheless, rater reliability has not been examined and the possibility of severity on the part of the professor–researchers should not be overlooked. Another potential explanation for this difference in scores could be that the students, many of whom have a high B2 or C1 level, feel sufficiently comfortable with their English and are, hence, able to talk into the computer with ease. Yet, the video recordings show that many of them appeared to be very relaxed during the initial interviews and even laughed with the researchers asking the questions, suggesting that the face-to-face situation did not involve any barriers to demonstrating their level accurately.

Degree – Education or English			Frequency	Percentage	Valid percentage	Accumulated percentage
Education	**Valid values**	B1-	4	8.3	9.3	9.3
		B1	8	16.7	18.6	27.9
		B1+	8	16.7	18.6	46.5
		B2-	11	22.9	25.6	72.1
		B2	6	12.5	14.0	86.0
		B2+	2	4.2	4.7	90.7
		C	4	8.3	9.3	100.0
		Total	**43**	**89.6**	**100.0**	
	Missing values		5	10.4		
	Overall total		**48**	**100.0**		
English	**Valid values**	B1	1	3.3	3.6	3.6
		B2-	3	10.0	10.7	14.3
		B2	6	20.0	21.4	35.7
		B2+	4	13.3	14.3	50.0
		C	14	46.7	50.0	100.0
		Total	**28**	**93.3**	**100.0**	
	Missing values		2	6.7		
	Overall total		**30**	**100.0**		

Table 4: Aptis Speaking Test score levels according to degree program

Results by faculty		N
Education	Aptis score < Face-to-face score	21
	Aptis score > Face-to-face score	16
	Aptis score = Face-to-face score (No difference)	6
	Total	**43**
Arts and Humanities	Aptis score < Face-to-face score	5
	Aptis score > Face-to-face score	15
	Aptis score = Face-to-face score (No difference)	8
	Total	**28**

Table 5: Comparison of university professors' scores and Aptis scores by faculty group

The results for students at the Faculty of Education, which revealed that the scores were higher for the face-to-face exercise, could have been related to their lower overall level of English, which was generally lower than that of the Faculty of Arts and Humanities' students. The former students may have felt shyer than the latter ones because of their lower levels and their more limited exposure to English both in and out of class. These students have taken far fewer university courses in English compared to the English Studies and Translation students, who take nearly all of their classes in the target language. In this sense, the chance to be evaluated in person and in the company of a fellow student in the same situation may have enabled the students to feel more at ease during the task.

The limitations of the comparison should also be taken into account in considering the results of both of the groups. The university researchers–raters and Aptis raters gave scores for two different exercises with different conditions, not simply the difference between an in-person task and a computer-delivered one. While the tasks in each were similar, the pressure related to each situation was different; the initial interview took place in the professors' offices as part of a research project that would involve no official stakes, while the Aptis Test took place in a computer laboratory that many, if not all, of the Education students had never seen prior to the testing date and under official exam conditions with the potential for a level certificate. A difference in timing between the interviews and the computer-delivered test does not appear to have been a factor as the students interviewed in November 2015 who completed the study belonged to the Faculty of Education group, which generally obtained higher scores on the initial face-to-face interview. A more reliable comparison to determine the situation of the university professors compared to the Aptis raters would need to involve the same exercises rated by the different raters. In this sense, the mock speaking test delivered prior to the official Aptis Test could serve, so the research team plans to rate the students' performance on this exercise in the future to compare their scores with those provided by the Aptis Test raters.

5.2.   Guided practice and repetition of the Aptis Test

The second research question involved determining whether knowledge of the different tasks in the Aptis Speaking Test and the opportunity to prepare for the test affected the students' results. Section 5.2.1 discusses the first part of this question while Section 5.2.2 relates to the second part.

5.2.1. Test scores

The scores for the different sections of the Aptis Test for the various testing dates are listed in Table 6. The members of the experimental group who took the test on 1 April obtained slightly lower scores overall than their peers in the control group. The opposite situation is observed for 8 April, when the experimental group members scored slightly higher than those in the control group. No comparison can be made between the two groups for the test-takers on 29 April as all of those participants were from the same group. However, a comparison of the scores for the experimental group on the three different testing days reveals an increase from 1 April to 8 April but a drop from 8 April to 29 April.

The most striking finding here is that the Speaking Test scores were consistently within the range of the other test scores for both the experimental and the control groups. In other words, the students' levels of speaking were similar to their levels in the other areas tested, unlike the discrepancies observed at University of Alcalá in recent years (see Table 7). At the same time, it was observed that the scores for the Listening Test were generally lower than the other areas. This result echoes the findings of the First European Survey on Language Competence (European Commission (2012), which showed that Spaniards obtain relatively lower scores in listening than people in other countries.

Of the 31 students who took the Aptis Test a second time, 11 students (more than 33% of the group) improved their speaking scores. This is a high number considering that the test was delivered in the same conditions and using similar prompts only three to four weeks after the first test date. This improvement was calculated by comparing the scores for the two Aptis Tests for the 25 students who had obtained scores in the more central range, as opposed to the outliers, with a view towards obtaining more sound statistics and inferences. The result can be considered to be very high, as it was the same type of exam.

	Experimental group					Control group				
	N	Min.	Max.	Mean	Std Dev.	N	Min.	Max.	Mean	Std Dev.
April 1 Grammar Vocabulary	11	27	41	33.91	5.576	23	26	50	38.22	5.768
April 1 Listening	11	26	44	35.82	5.173	23	24	50	38.87	6.518
April 1 Reading	11	26	48	40.55	6.933	23	26	50	43.57	6.089
April 1 Speaking	11	34	41	37.45	2.659	23	26	47	41.91	4.602
April 1 Writing	11	32	44	37.45	3.475	23	26	50	40.52	5.333
April 1 Total	**11**	**136**	**163**	**151.27**	**10.335**	**23**	**130**	**193**	**164.87**	**14.150**
April 8 Grammar Vocabulary	34	25	48	41.00	5.893	7	28	44	36.14	5.014
April 8 Listening	34	26	50	42.41	5.863	7	22	46	35.14	9.371
April 8 Reading	34	20	50	45.24	6.769	7	28	46	39.71	6.676
April 8 Speaking	34	28	50	44.50	4.724	7	34	48	39.57	5.192
April 8 Writing	34	32	48	43.47	4.158	7	32	44	39.14	4.298
April 8 Total	**34**	**109**	**198**	**175.62**	**18.671**	**7**	**122**	**179**	**153.57**	**20.631**
April 29 Grammar Vocabulary	24	27	50	40.00	5.786					
April 29 Listening	24	24	48	39.92	6.171					
April 29 Reading	24	20	50	44.83	7.124					
April 29 Speaking	24	26	48	41.42	5.508					
April 29 Writing	24	24	50	42.42	6.296					
April 29 Total	**24**	**107**	**196**	**168.58**	**21.847**					

Table 6: Score information for the two groups on the three test dates

Nevertheless, the increase in speaking scores was small, as was the overall sum for the experimental group. This suggests a tendency towards better scores that are not significant and implies that guided, mediated repetition of the test can lead to improved results, albeit small ones (Poehner, 2007).
It should be pointed out though that the research team believes that repetition itself with no further guidance or self-reflection is not likely to lead to better results. Improvement in this case could have taken place because usability improved and/or the students understood how the exam works and this favoured their performance. In any case, the largest improvement in scores was seen in reading and writing. The biggest increase in scores in general was found in the Faculty of Education, which tended to have scores at the B2 level.

	Mean	N	Standard deviation	Standard error of the mean
G/V Aptis 1	38.24	25	6.942	1.388
G/V Aptis 2	39.84	25	5.720	1.144
Listening Aptis 1	39.76	25	5.953	1.191
Listening Aptis 2	39.68	25	6.156	1.231
Reading Aptis 1	43.60	25	7.767	1.553
Reading Aptis 2	45.04	25	7.050	1.410
Speaking Aptis 1	**40.96**	**25**	**5.303**	**1.061**
Speaking Aptis 2	**41.20**	**25**	**5.500**	**1.100**
Writing Aptis 1	40.72	25	4.826	.965
Writing Aptis 2	42.56	25	6.205	1.241
Total Aptis 1	**165.04**	**25**	**20.643**	**4.129**
Total Aptis 2	**168.48**	**25**	**21.393**	**4.279**

Table 7: Improvement between Aptis 1 (1 or 8 April) and Aptis 2 (29 April)

5.2.2. Guided practice

As mentioned in Section 2.4.1 above, the students participated in a mock speaking test in the hour before they took the first Aptis Test. The purpose of this exercise was to explore whether a situation with a certain level of anxiety and the use of notes would lead to significantly different results from the actual Aptis Speaking Test. During the mock test, the students were only given the instruction that they could prepare for 15 minutes but no indications of how to do so were provided.

The scores for the mock test and first Aptis Speaking Test were compared overall to determine whether the students obtained different levels for the simulated situation compared to the actual test situation (Table 8). Both exercises were rated by Aptis evaluators, but the former were calculated out of a possible of 48 points while the latter were calculated out of a maximum of 50 points since the grammar and vocabulary score is taken into account in this case.

This first overall comparison indicates that the scores for the Aptis Speaking Test were higher than the scores for the mock test – a situation that could have been expected as the professors in the Department of Modern Philology at University of Alcalá tend to have the impression the Aptis is an easy test. The difference in maximum possible scores for the two formats can also account for some of this result. The difference in the mean scores, when translated into actual CEFR levels, however, actually represents two different levels: B2 (41.98 points) and B1 (34.23 points).

	N	Range	Min	Max	Mean	Std Dev.	Var.	Asym	Curtosis
Aptis 1 Speaking	84	24	26	50	41.98	5.082	25.831	-.786	.397
Mock test	87	24	20	44	34.24	5.618	31.557	-.255	-.397

N valid = 80 (students who did both Aptis 1 and mock test)

Table 8: Aptis Speaking Test 1 and Mock speaking test comparison

A direct comparison could be made of the scores for 80 students who completed both the first Aptis Test and the mock test. In order to determine if there were any differences related to relatively higher or lower scores, these 80 pairs of scores were divided into groups by quartiles, with the second and third quartile included in a middle group (Table 9). A total of 15 scores fell into the lowest quartile, 43 scores belonged to the middle group and 24 scores fit into the highest one. It was found that all three groups obtained higher scores on the Aptis Test as opposed to the mock test. It was also observed that the lowest group showed the most variation (9.133) and the middle group showed the next largest variation (8.372).

Mock test pairs		Mean	N	Std deviation	Standard error of the mean
Low	Pre-test Speaking	35.27	15	3.327	.859
	Mock speaking test	26.13	15	3.502	.904
Middle	Pre-test Speaking	41.81	43	3.996	.609
	Mock speaking test	33.44	43	2.657	.405
High	Pre-test Speaking	46.45	22	2.444	.521
	Mock speaking test	41.18	22	1.181	.252

Table 9: Aptis Speaking Test and mock speaking test by performance group

These results might be explained by the fact that many of the students were observed to write out their complete answers during the preparation time prior to the mock exam. This situation lent itself naturally to their reading of their responses instead of conversing spontaneously when it was time to record their answers. Responses that are read can be expected to have less natural intonation and pronunciation, which could easily explain why the scores to the mock test were lower. In addition, a number of the students did not have time to finish writing out answers to the final questions and, hence, had to improvise those responses, another factor which may have led to lower scores. Despite the lower results, this practice session may have helped the students to improve their scores, since it assisted them in becoming familiar with the exam format and its difficulty.

5.3.   Student self-perception and Speaking Test scores

After the students viewed the recordings of their initial interviews, they responded to a questionnaire (Appendix 3) on their perceptions of their performance. The results for the 11 students who obtained higher scores on the second official Aptis Test show that they tended to rate themselves close to the B2 level (Table 10).

	Level							
	A2		B1		B2		C1	
	F.	%	F.	%	F.	%	F.	%
The degree of fluency in your video is most similar to which model video?	1	9.1	5	45.5	3	27.3	2	18.2
The pronunciation in your video is closest to which model video?	1	9.1	3	27.3	5	45.5	2	18.2
Your ability to communicate ideas in your video is closest to which model video?	1	9.1	4	36.4	4	36.4	2	18.2
The level of grammar that you used in your video is closest to which model video?	1	9.1	2	18.2	6	54.5	2	18.2
The level of vocabulary that you used in your video is closest to which model video?	1	9.1	5	45.5	3	27.3	2	18.2
The length of the comments that you make in your video corresponds most closely to which of the model videos? (number of words per response/length of each sentence)	1	9.1	3	27.3	4	36.4	3	27.3
The way of connecting ideas in your video is closest to which model video?	1	9.1	3	27.3	4	36.4	2	18.2
Which overall level would you give yourself based on the model videos?	1	9.1	4	36.4	4	36.4	2	18.2

Table 10: Student self-evaluations of their initial interview performance

Only one of these students identified his performance at the A2 level, while two of them rated themselves at C1; the other eight students put themselves into the B band. The data in Table 8 reveal that the students saw themselves as having pronunciation, vocabulary and grammar that were adequate for the level, but that they often gave themselves a comparatively lower mark for fluency. It must be highlighted, however, that this group of students tended to have higher Aptis scores for grammar and vocabulary. The questions in the questionnaire on the length of their comments and the connection of ideas may have been too difficult for the students to answer as they had not received any prior explanations; nevertheless, the responses were in line with the rest of the answers as they followed the same trend towards a B level.

In response to a qualitative question on the same questionnaire that indicated to the students that they could add any additional information about their level, the participants provided a number of responses (translated by the authors from Spanish).

- After listening to it now I realise what mistakes I made and I know how I should have said it, but at the time it is really difficulty to speak.

- I think that I speak better on some occasions, like when I explain my experience abroad, compared to when I describe a photograph.

- In cases like the description of photographs, I see that I am less fluent because I have to improvise a description based on photos that don't make me think of anything.

- In my case, it is really stressful to be recorded while doing an oral exam in English. I think I could have done it better and I am upset with the result. I made grammatical mistakes that cannot be excused and I don't understand why.

- First of all, I must say that I was really nervous so I forgot a lot of basic words. In addition, I sometimes need to think a little more in order to build sentences. In addition, when I really feel more comfortable and relaxed, I enjoy myself more and don't think so much, I am more spontaneous.

- The fluency in the first questions or at the beginning of the recording is not the same since I was more nervous at the start.

- My nerves always get the best of me! I could have done it better but I do need more vocabulary. I also must admit that I am very hard on myself.

- It was difficult for me to choose between levels B1 and B2 since I could see differences between the two people in the videos. Depending on the person I focused on in the video, it could have been one level or the other. I focused on the girl in the B1 video who had a higher level.

- I can see insecurity and nerves.

As can be seen in the results, nerves are of major importance in the responses of the students who feel that they could have performed better in a less stressful situation. Some of them even point out that they do better in situations that do not involve interaction with a person. In addition, some students were able to perceive their weaknesses and improve in the second round of the Aptis Test. Nevertheless, although data on the participants' perceptions is provided through these comments, no follow-up on the effect was done.

5.4. Aptis Speaking score and the other skills scores

As mentioned in Section 4.2.1 above, the students studied here tended to obtain speaking scores that were similar to the other sections of the Aptis Test. This situation is in striking contrast to that observed in earlier years by professors at University of Alcalá, when students were found to score as many as two levels lower on the Speaking Test compared to the other skills. However, this could be due to a contextual matter. Since most students who take the Aptis Test come from the Faculties of Education and Arts and Humanities, and about 80% of the Aptis sessions are delivered in the latter, the perspective could be biased by the fact that the main local informants (professors at Arts and Humanities) observe that their "around C" students perform better on the Aptis Test than they do in face-to-face exams.

It is also important to highlight here that the scores for the Listening Test obtained by the students in this study tended to be consistently lower than the other parts of the test, a situation not perceived to date by the professors at the faculty. Nevertheless, as mentioned earlier, this result is consistent with the findings of the European Survey of Language Competence, which have shown that Spaniards tend to be weak in this language skill.

6. DISCUSSION

As described above, it has been found that the point of view of the professors at University of Alcalá regarding the Aptis Test is clearly different from the results of this study. While there is a common complaint that the test does not reflect students' knowledge accurately, the data obtained here reveal some highly interesting findings.

1. The differences between the levels assigned by the University of Alcalá professors for the face-to-face initial oral interviews and the official Aptis Test results for the students at the C1 level are not significant.

2. In comparing the results, a third of the time (33.3%), the Aptis Test result is lower than the score assigned for the initial interview. However, almost half of the time (47.43%), the Aptis Test results are higher than the score given by the University of Alcalá raters. This difference is not significant but it is high enough for the professors in general to notice it and to have the impression that the Aptis Test is easy.

3. It is interesting to note the percentage of evaluations that coincided (19%) because the University of Alcalá raters did not have access to the Aptis evaluation criteria and rated the students using different rubrics. In this sense, it can be said that this result is positive. However, at the same time, our overall results suggest that the Aptis Test may benefit from a larger external validation study to compare it to other oral exams such as the Cambridge Suite or the IELTS test.

4. In terms of the different groups studied here, the students at the Faculty of Education obtained slightly better results on the initial face-to-face interview than on the Aptis Test but the difference is not significant. However, the results for the Faculty of Arts and Humanities' students were three times better on the Aptis than in the initial interviews (53% versus 17.8% in the best cases).

It was expected that the students would do better on the Aptis Test than on the mock speaking test because of the discrepancies in levels observed by professors compared to the Aptis Test in recent years. The overall difference between the two formats came to 7.66 points in favour of the official test, a substantial difference considering that it is enough in many cases for a student to reach the next level. The middle and lower level groups improved the most with an increase of approximately 9 points on the official test. These differences in student level are especially important and can be related to three possible causes:

- greater coherence and fluency in the responses
- better vocabulary
- greater attention to language forms.

In any case, if these are the only characteristics that improve between the mock test and the official test, it should not be enough to cause a candidate to move up a level. However, many students in this study who obtained a C1 on the Aptis Speaking Test obtained a B2 for the mock test when it was evaluated by the Aptis raters. This improvement can be seen in the high correlation in the paired sample statistics. The Aptis raters were most likely aware that the students had limited themselves to reading their prepared responses and that some of them had run out of time to prepare in the majority of the cases.

Other reasons to explain the increase in Aptis Test scores could be the fact that the students had become familiar with the test format or they had relaxed as a result of the practice they had obtained doing the mock test. In addition, the students' levels and the potential for improvement could have been a factor involved in the score differences.

The students at a C1 level showed less improvement on the official test possibly because the Aptis Test is designed to evaluate candidates up to the level of C. This would mean that the lower level students had more room for improvement. Nevertheless, a more extensive study with a larger population would be of interest to obtain better inferences.

7. CONCLUSION

A number of findings that are not statistically significant have been observed and are worthy of further study with a larger population of subjects. First, there was some degree of discrepancy in the scores assigned to students for their speaking performance in a face-to-face situation by professors at the University of Alcalá and the Aptis Speaking Test scores that the students obtained a few weeks later. In fact, one of the groups in the study scored better on the Aptis Tests three times more often than they did on the face-to-face interview situation. The university professors who took part in the study have extensive experience in testing students for their levels, but it must be highlighted that no inter- or intra-rater reliability checks were carried out. In order to determine how a face-to-face speaking test situation compares to a computer-delivered one, it would be more effective to compare student test scores on two official speaking tests, the Aptis Test and an IELTS or Cambridge Test, for example.

Second, it seems that knowledge of the different tasks on the Aptis Test and repetition of the test shortly after a first test date leads to a slight improvement in scores. This is especially the case for students at the B2 level as opposed to the C1 level. Increases were observed for speaking and especially in reading and writing. Listening, a weak area for Spaniards, also improved but it was found consistently to be the lowest of all the scores among the students participating in this study. It would be of interest to examine the effect of students' relatively lower levels of listening ability on their capacity to perform effectively on a speaking test. In the case of a computer-delivered test that includes written prompts on the screen, it might be less significant but this factor could have an impact in other tests that involve oral comprehension during the speaking section of the test.

Third, a subgroup of students in this study who obtained higher scores on a second Aptis Test tended to perceive their overall speaking level to be in the B range. In particular, they rated their own pronunciation and grammar in a face-to-face interview situation as being in the B2 level, while they tended to place their fluency and vocabulary in the B1 level. However, their actual grammar and vocabulary scores on the Aptis Test were higher than their own perceptions of their performance. A number of comments made by the students in terms of their impressions of their face-to-face interviews revealed that they were highly critical of their performance and that they felt nervous due to the fact that the interview was being recorded. In this sense, it can be said that they were aware of their limitations and felt that they could have performed better in more relaxed circumstances.

Finally, no difference was found between the Speaking Test scores and the scores for the other parts of the Aptis Test, unlike the situation that has occurred on repeated occasions in the past at the University of Alcalá.

REFERENCES

European Commission, Language and Training. (2012). *First European Survey on Language Competences*. Retrieved on 11/11/2016 from http://ec.europa.eu/dgs/education_culture/ repository/languages/policy/strategic-framework/documents/language-survey-final-report_en.pdf

He, L. and Shi, L. (2008). ESL Students' Perceptions and Experiences of Standardized English Writing Tests, *Assessing Writing*: 13 (2) 130–149.

Oller, J. (1979). *Language Tests at School: A Pragmatic Approach*. London: Longman

O'Sullivan, B. (2012). Aptis Test Development Approach, Aptis Technical Report. London: British Council. https://www.britishcouncil.org/sites/default/files/aptis-test-dev-approach-report.pdf

Poehner, M. (2008). *Dynamic Assessment A Vygotskian Approach to Understanding and Promoting L2 Development*. Berlin: Springer.

Simpson, J. (2006). Differing Expectations in the Assessment of the Speaking Skills of ESOL Learners. *Linguistics and Education* 17 (1): 40–55.

Van Moere, A., Suzuki, M., Downey, R. and Cheng, J. (2009). Implementing ICAO Language Proficiency Requirements in the Versant Aviation English Test. *Australian Review of Applied Linguistics* 32 (3): 27.1–27–17.

Weir, C. (2005). *Language testing and validation: An evidence-based approach*, Basingstoke: Palgrave MacMillan.

Xie, Q. and Andrews, S. (2013). Do Test Design and Uses Influence Test Preparation? Testing a Model of Washback with Structural Equation Modelling, *Language Testing* 30 (1): 49–70.

APPENDIX 1:
The Pre-test Interview Prompts

Task 1 (30 seconds): Questions were taken from the list below, and mainly from the first half of the page. The students worked individually.

Where do you live?
How long have you lived there?
Who do you live with?
Would you say it is a good place to live?
What is the best thing about where you live?

Where are you from?
When did you leave XXXX?
What do you like about XXX?
What kind of jobs do people in XXX have?
How does it compare to Alcala?

What do you do in your free time?
How did you learn to do XXX?
Who do you do XXX with?
What do you like about XXX?
What would I need to do to learn to do XXX?

What did you do this morning (yesterday
    afternoon/last weekend)?
What are you going to do after this test
    session?

Have you ever been abroad/to another
    country?
When was it?

What did you do there?
Do you recommend it?
How does it compare to Spain/Alcala?
What are you studying?
Why did you decide to study it?
What do you want to do when you finish your
    studies?
What are the advantages to studying this
    program?

Who is your best friend?
Where did you meet?
How often do you see each other?
What kinds of things do you do together?
What is his/her best characteristic?

When is your birthday?
How do you normally celebrate it?
What is your opinion of celebrating birthdays?
When was your best birthday?

Do you drive?
Do you like it?
How often do you drive? What for?
Would you like to learn?
What kind of car would you like to have? Why?

Task 2 (45 seconds): The students were asked to describe a photo individually. The photos were taken from the series of books of past FCE papers from prior to the revised exam published by Cambridge.

Task 3 (45 seconds): The students were asked to work on their own to compare two photographs from the same bank of photos.

Task 4 (one minute for preparation followed by two minutes to speak): The students prepared a series of questions from the following list and then gave a monologue:

Tell me about a time you won a prize or an award.
How did you feel about it?
Do prizes encourage people to do their best?

Tell me about one of the most exciting situations / times in your life.
How did you feel?
Did the situation affect your life?

Describe the best concert of sport match you have attended.
Why was this the best?
Are athletes, actors or musicians paid too much? Why?

Is being a teacher a difficult job?
How do you feel about it?
Do you think your studies at university are good enough to prepare you?

Task 5 (45 seconds): One student was asked to describe a photograph from the bank of pictures, while the other listened and then asked two questions. Then the roles were reversed for a second picture.

Task 6 (one minute for preparation followed by a two minute dialogue): The students were given a role play such as the following and worked together in doing it.

Situation 1
Student A: You are a student at university. You are not happy with the mark you received on your last exam. Talk to your professor and try to convince him/her to give you a higher mark.

Student B: You are a university professor. Your student wants a higher mark and has come to see you in order to ask for a higher mark. Decide if it is possible and under what conditions.

Situation 2
Student A: You are talking to your English boyfriend because you are not happy with the way he acts with you. Describe the problem to him and try to find a solution to the problem with him.

Student B: You are an English boy. Your Spanish girlfriend is not happy. Talk to her to see if there is a solution to the problem.

Situation 3
Student A: You are a Spanish student in London for four months. You are meeting a person who has a room to rent. You want to see if the two of you are compatible as flat mates.

Student B: You are an English student in London and you have an extra room to rent. You are going to talk to a person who wants to rent your room. Talk to him/her to see if you are compatible.

Situation 4
Student A: You are a student who lives with his/her parents. Your parents insist that you must come home at 11.00 on Friday/Saturday nights but you want to stay out later. Discuss the problem with your parents.

Student B: You are the mother/father of a 20-year-old girl/boy. She/he likes to stay out late on Friday and Saturday nights. Discuss the situation with him/her. What solution can you find?

APPENDIX 2: Questionnaire 1

Please answer the following questions by writing the answer or selecting the most appropriate option with an X

1. Student code number (write the number assigned to you by the teacher)	_____
2. Age	1) 18–20 2) 21–22 3) 23–25 4) More than 25
3. Sex	1. Male 2. Female
4. Program of study	1. Early Childhood Education 2. Primary Education 3. Modern Languages and Translation 4. English Studies
5. Year of study	1. First 2. Second 3. Third 4. Fourth
6. Indicate the number of years that you studied English before starting your university studies	1. 5-7 years 2. 7-12 years 3. More than 12 years
7. How many subjects have you studied in English until today (include last semester) (answer this question only if you are studying at the Faculty of Education)	1. 1 2. 2-3 3. More than 3
If you have studied English outside of your classes to date, how have you done so? You may select more than one option.	1. At a private language school 2. With an private teacher 3. On my own 4. By spending time in an English-speaking country 5. Others (indicate what you have done)
9. Have you ever done any kind of stay in another country?	1. Yes 2. No
9.1. If you answered yes, how much time did you spend in the other country? If you have done more than one stay in another country, you can indicate more than one response.	1. Less than a week long 2. One week 3. 10 days-15 days 4. Three weeks 5. One month 6. More than one month (indicate length) _____
10. How often have you used English up to now?	1. Never or almost never 2. A little, only out of need (in class, at work). 3. Sometimes (in class, at work, to talk to people sometimes out of a classroom situation) 4. Normally, I speak English in class as well as outside of class. 5. Often, I take advantage of any occasion to be able to speak English.
11. Have you ever taken the Aptis Test at any time before participating in this project?	1. Yes 2. No
11.1. If you answered yes, when was the last time you took the Aptis Test?	1. In the last month 2. In the last six months 3. Between six months ago and a year ago 4. Between one year and two years ago 5. More than two years ago
12. Have you ever taken a specific course to prepare for an official English exam?	1. Yes 2. No
12.1. If you answered yes, how did you prepare for the official English test?	1. At a school 2. Another way (specify it)

12.2.	If you answered yes, which official English test did you prepare? You may select more than one choice.	1. KET 2. PET 3. FCE (First) 4. CAE (Advanced) 5. EOI 6. TOEFL 7. TOEIC 8. IELTS 9. Another one specify it _____
13.	Have you ever been recorded when speaking in English?	1. Yes 2. No
13.1.	Have you ever seen or heard yourself in a (video) recording in English?	6. Yes 7. No
13.2.	If you answered yes, did you feel uncomfortable being recorded while speaking in English during the interview?	1. Yes 2. No If you answered yes, indicate why:
14.	Have you ever taken an official English test on a computer?	1. Never 2. 1-2 times 3. A few times (3-5 times) 4. More than 5 times
14.1	If you answered yes, when was the last time you took an English test on a computer?	1. In the last month 2. In the last 6 months 3. Between 6 months and a year ago 4. Between one year ago and two years ago 5. More than 2 years ago
15.	Have you ever taken another kind of test on a computer?	1. Never 2. 1-2 times 3. A few times (3-5 times) 4. More than 5 times
15.1.	If you answered yes, indicate which test you took on a computer.	

Questions about the oral interview with the teacher

Please indicate the extent to which you agree with the following statements following a scale of 1 to 5; a value of 1 indicates that you disagree completely while a value of 5 means that you agree completely.

1	2	3	4	5
Disagree completely	-	-+	+	**Agree completely**

Questions	Answers				
16. I am satisfied with the level that I showed during the interview	1	2	3	4	5
17. I think that my responses correctly show my level of English	1	2	3	4	5
18. I have level B2 for oral expression (Speaking) in English	1	2	3	4	5
19. I have a level B2 for oral comprehension (Listening) in English	1	2	3	4	5
20. I have a level B2 for written comprehension (Reading) in English	1	2	3	4	5
21. I have a level B2 for written expression (Writing) in English	1	2	3	4	5
22. I am capable of reaching a level of B2/C1 on an official test in the future.	1	2	3	4	5
23. I felt nervous with the interviewer	1	2	3	4	5
24. I felt nervous with the video camera	1	2	3	4	5
25. I felt comfortable talking with my partner during the interview	1	2	3	4	5
26. I understood what I had to do at all times during the interview	1	2	3	4	5
27. I understood the questions that I was asked at all times	1	2	3	4	5

28.	I was able to speak well during the monologues	1	2	3	4	5
29.	I was able to speak well during the dialogues	1	2	3	4	5
30.	I had enough time to prepare for the parts of the interview that allowed a minute to write down ideas	1	2	3	4	5
31.	My notes helped me to answer the questions when I had to speak	1	2	3	4	5
32.	I take effective notes in my classes that are taught in Spanish	1	2	3	4	5
33.	I take effective notes in my classes that are taught in English	1	2	3	4	5
34.	I have been taught in class how to take notes	1	2	3	4	5
35.	I have been taught how to take notes effectively	1	2	3	4	5

36. I think that my level of English as demonstrated in the interview is level…	1. A1 2. A2 3. B1 4. B2 5. C1 6. C2

Open-ended questions

37. Describe in as much detail as possible how you think a person with a B1 level in English speaks.

38. Describe in as much detail as possible how you think a person with a B2 level in English speaks. What is the difference between a person who speaks at the B1 level and one who speaks at the B2 level?

39. Describe what you know about the Aptis Speaking Test format.

40. What strategies will you need to improve your level on the Aptis Test? Describe them.

APPENDIX 3: Questionnaire 2

Please listen to the videos with models of the levels of English provided by the teachers and answer the following questions.

Questions	Answers					
1. The degree of *fluency* in your video is most similar to which model video?	A1	A2	B1	B2	C1	C2
2. The *pronunciation* in your video is closest to which model video?	A1	A2	B1	B2	C1	C2
3. Your *ability to communicate ideas* in your video is closest to which model video?	A1	A2	B1	B2	C1	C2
4. The level of *grammar* that you used in your video is closest to which model video?	A1	A2	B1	B2	C1	C2
5. The level of vocabulary that you used in your video is closest to which model video?	A1	A2	B1	B2	C1	C2
6. The length of the comments that you make in your video corresponds most closely to which of the model videos? (number of words per response/length of each sentence)	A1	A2	B1	B2	C1	C2
7. The way of connecting ideas in your video is closest to which model video?	A1	A2	B1	B2	C1	C2
8. Which overall level would you give yourself based on the model videos?	A1	A2	B1	B2	C1	C2

9.   Indicate any other ideas you would like to add about your answers to the above questions.

10. Have you ever received help in understanding what each level involves? When and where? What information did you receive?

11. Have you taken an Aptis Test before? Do you remember the levels that you obtained?
Speaking   Writing   Listening comprehension   Reading

12. Now that you have heard your interview, what can you do to reach a higher level?

13. What difficulties do you have when being recorded or video recorded during an oral test? How do you feel about anxiety, mistakes and other feelings?

14. Besides passing the test, do you have any other objectives when you have to take an oral exam (for example, pronounce well, improve the grammar, etc.)?

15. Do you try to correct yourself or do you simply think that the way your ideas come out will be best? If you correct yourself, how much do you think you do so?

16. How do you help yourself to calm down/relax before an oral interview? Or do you not need to do so? Or do you not know how to do so?

17. What do you focus on when you take notes?

18. What strategies do you use during an individual oral exam? And during one done in pairs?

19. What do you do or think if your partner does not respond or interact?

20. How do you facilitate interaction with your partner?

21. Is it important to know your partner? Why? How does it help you?

APPENDIX 4:
Mock Exam Handout

Practice exercise prior to the official Aptis Test (Speaking)

You will have 15 minutes to practice responses to the following questions.
Afterwards, Mary Frances will read the questions and you will record your answers on the computer.
(Instructions translated from Spanish)

Part 1
    0. Say your full name. Be sure to speak loud and clear.
    1. Tell me about your family (30 seconds)
    2. Describe your typical Saturday (30 seconds)
    3. What is the weather like today (30 seconds)

Part 2
    4. Describe this picture (45 seconds)

    5. Why is it important for young people to live on their own and not with their families?
       (45 seconds)
    6. Describe the place where you live. (45 seconds)

Part 3

7. What do you see in the two pictures below? (45 seconds)

8. Which form of studying works best for you? Alone or in a group? Why? (45 seconds)
9. What would you do if you had to work in the other way often? (45 seconds)

10. Answer the next three questions together (2 minutes for all of them)

    a. Tell me about a time that you participated in a sport.
    b. How did you feel about it?
    c. Do you think children should be encouraged to participate in sports on a regular basis?

APPENDIX 5:
Questionnaire 3 – Answered Online

Please answer the following questions. Some of them require you to write a response while others involve selecting a number from one to four. In the latter cases, a 1 means that you are "not at all in agreement" or "never." On the contrary, a 4 means that you are "completely in agreement" or "always." Please respond to the questions before Friday, 22 April, at 24.00. Thank you very much.

Full name

1. What differences did you find between the interview and the computer-based test?

2. Which of the two formats do you feel more secure using? Why?

3. Which type of test (a test with an interview or a computer-based test) do you think is better to evaluate your real ability in spoken English? Why?

4. Which way of taking a test is better for you? Why?

5. Is there anything that makes you feel uncomfortable or insecure?

6. I used the material that the researchers gave me. (If they gave me materials)
   (Leave the question blank if you did not receive material).
   Totally disagree / Never   1   2   3   4   5   Totally agree/Always

7. Now that I have taken the official Aptis Test, I can say that the materials really helped me.
   Totally disagree / Never   1   2   3   4   5   Totally agree/Always

8. If I did not receive any support material from the researchers, I made an effort to prepare for the test.
   Totally disagree / Never   1   2   3   4   5   Totally agree/Always
   If you answered yes, what did you do?

9. The computer is an appropriate method for the Aptis Speaking test.
   Totally disagree / Never   1   2   3   4   5   Totally agree/Always
   Why?

9. The computer measures my spoken ability in English.
   Totally disagree / Never   1   2   3   4   5   Totally agree/Always
   Why?

10. The simulation in the hour before the official Aptis tests helped me during the actual exam itself.
    Totally disagree / Never   1   2   3   4   5   Totally agree/Always

11. Taking notes before the recorded simulation helped me to improve my response during t he simulation itself.
    Totally disagree / Never   1   2   3   4   5   Totally agree/Always
    In what ways did taking notes help you most?

13. Taking notes during the official Aptis Speaking test helped me to improve my response during the test itself.
Totally disagree / Never   1   2   3   4   5   Totally agree/Always
How did your notes help you?

14. I read my notes during the simulation (in other words, I did not speak freely).
Totally disagree / Never   1   2   3   4   5   Totally agree/Always

15. I read the notes that I prepared during the Aptis Speaking Test when it was time to respond to the speaking tasks.
Totally disagree / Never   1   2   3   4   5   Totally agree/Always

16. This way of preparing for the Aptis Speaking Test (trial with a person, recorded trial, official Aptis Speaking Test) helped me to obtain a good result.
Totally disagree / Never   1   2   3   4   5   Totally agree/Always
How? In what sense?

17. The Aptis application was reasonable to handle.
Totally disagree / Never   1   2   3   4   5   Totally agree/Always

18. It is easy to see how the Aptis application works.
Totally disagree / Never   1   2   3   4   5   Totally agree/Always

19. The setup of the self-evaluation session was adequate.
(If you did not attend it, leave this question blank.)
Totally disagree / Never   1   2   3   4   5   Totally agree/Always

20. The self-assessment session helped me to determine my actual level of spoken English.
(If you did not attend it, leave this question blank.)
Totally disagree / Never   1   2   3   4   5   Totally agree/Always

21. As a result of the self-assessment session, I stated to think about strategies for improving my performance. (If you did not attend it, leave this question blank.)
Totally disagree / Never   1   2   3   4   5   Totally agree/Always

British Council
Assessment Research
Awards and Grants

If you're involved or work in research into assessment, then the British Council Assessment Research Awards and Grants might interest you.

These awards recognise achievement and innovation within the field of language assessment and form part of the British Council's extensive support of research activities across the world.

STUDENT PERCEPTIONS OF THE CEFR LEVELS AND THE IMPACT OF GUIDED PRACTICE ON APTIS ORAL TEST PERFORMANCE

Jesús García Laborda,
Marián Amengual Pizarro,
Mary Frances Litzler, Soraya García
Esteban, Nuria Otero de Juan