UNIVERSITY OF MURCIA

International Journal of English Studies

IJES

http://revistas.um.es/ijes

How are spoken skills assessed in proficiency tests of general English as a Foreign Language? A preliminary survey[1]

Mª LUISA ROCA-VARELA & IGNACIO M. PALACIOS*
University of Santiago de Compostela

Received: 8 June 2013 / Accepted: 4 October 2013

ABSTRACT
This paper examines some of the best known proficiency tests in English, with particular focus on the oral component. Attention is paid to the following issues, among others: the weighting of oral elements in testing, the criteria used for the assessment of oral skills and the relation of these to the general guidelines in the *Common European Framework of Reference* (CEFR), and the kinds of tasks and marking systems used for assessment. Our aim is to evaluate these tests as a means of determining the extent to which they can be considered valid tools for the assessment of oral performance, considering their significance, relevance and implications in the broader context of the modern world. We contend that, as is the case with teaching processes in general, which need to be continuously evaluated and reformulated, this is true of English testing, especially oral skills, given the special nature of the spoken language.

KEYWORDS: testing, spoken language, spoken tests, general English tests, assessment criteria, spoken tasks, CEFR

RESUMEN
Este artículo presenta un análisis de los exámenes oficiales de inglés más reconocidos a nivel internacional con atención especial a sus componentes orales. Se estudian, entre otros, los aspectos siguientes: la importancia de la parte oral en estos exámenes, los criterios de evaluación utilizados para las destrezas orales y su relación con las directrices emanadas del *Marco Común Europeo de Referencia*, las actividades orales que se incluyen en estos tests y el sistema de valoración utilizado. Nuestro objetivo último en la evaluación de estos exámenes es comprobar en qué medida se pueden considerar como instrumentos válidos para la evaluación del inglés oral, teniendo en cuenta la importancia, relevancia e implicaciones de estos exámenes en nuestra sociedad. Defendemos que, así como el proceso de enseñanza requiere una continua evaluación y reformulación, lo mismo debería aplicarse a los exámenes generales de inglés con especial atención a los componentes orales dadas las especiales características de la lengua hablada.

PALABRAS CLAVE: pruebas, lengua oral, exámenes orales, pruebas generales de inglés, criterios de evaluación, tareas orales, MCER

Address for correspondence: Mª Luisa Roca-Varela, Dpto. Filoloxia Inglesa, Facultade de Filoloxia, Universidade de Santiago de Compostela, Avda. de Castelao, s/n, 15782 Santiago de Compostela, Spain. Email: luissa10@yahoo.com Ignacio M. Palacios, Dpto. Filoloxia Inglesa, Facultade de Filoloxia, Despacho 414, Universidade de Santiago de Compostela, Avda. de Castelao, s/n, 15782 Santiago de Compostela, Spain. Email: ignacio.palacios@usc.es

Print ISSN: 1578-7044; Online ISSN: 1989-6131

1. INTRODUCTION

The status of English as the language of international communication (Mauranen & Ranta, 2009) has led to many people learning English as a foreign or second language in order to improve their career prospects, to travel, or to gain professional experience abroad. It is sometimes not enough to say that you can speak English, and a valid and internationally recognised certificate is required which states the exact level of your proficiency. Indeed, evidence of English proficiency has become a requirement for the admission of non-native speakers in most colleges and universities in the UK, Canada, Australia, New Zealand and the US, as well as in many other academic institutions around the world. The same applies to many companies and organisations which also require from candidates a particular language certificate for employment.

A wide range of English tests are available, including Cambridge ESOL, IELTS and TOEFL, all of which measure the learners' ability to communicate in English and also provide specific information on the mastery of the traditional four language skills: listening, reading, writing and speaking. Of those skills, oral production is commonly said to be the most complex ability to test, due to its specific features, the long time required for its assessment and the transient nature of the speech act. Although some language testing systems record candidates' production in order to have a permanent record of their spoken performance, speaking skills are still difficult to assess. As Heaton (1990: 67-68) claims, "whatever system is adopted, the marking itself is very subjective. We must take care, for example, to avoid allowing a student's personality to influence the grade we award". In addition to this, it is not easy to establish a definite list of the spoken features to be assessed, the specific criteria that are going to be considered for that assessment, or the activities or tasks to be used. Furthermore, if compared with the testing of other language skills, the testing of the oral proficiency, especially when done at large scale, can be regarded as the most difficult to carry out in terms of the test design and general organisation.

This paper examines some of the most popular standard English tests and their main features, paying particular attention to their oral modules. We focus specifically on the weighting of the oral part, the criteria used for the assessment of oral skills, and their relationship with the general guidelines provided by the Council of Europe (2001) in the *Common European Framework of Reference* (CEFR), the tasks used for assessment, and the marking system employed. Our aim is to evaluate these long established and widely recognised tests to determine the extent to which they can be regarded as valid instruments for the assessment of oral performance, considering their significance, relevance and their implications in the modern world. We believe that, as is the case with teaching processes in general, which need to be continuously evaluated and reformulated, the same should apply to existing official tests of English, with particular attention to their spoken oral modules.

Print ISSN: 1578-7044; Online ISSN: 1989-6131

2. ENGLISH LANGUAGE TESTS WORLDWIDE: AN OVERVIEW

As noted above, many international institutions now ask their prospective students, workers, researchers or language assistants to certify their level of English by means of widely recognised certificates, such as those of Cambridge ESOL, IELTS or TOEFL. It is also not uncommon for universities to design and conduct their own testing systems, in order to satisfy the growing demands of their students, who often need to justify their command of English. This is currently the case in many Spanish universities, where, as a consequence of the Bologna Declaration on the European Space of Higher Education (ESHE), undergraduates and sometimes also graduate students need to show that they possess at least a B1 level of a foreign language, in most cases English, when they graduate.[2] It is believed that a good performance on these tests will ensure the ability of these non-native speakers of English to communicate effectively in an English environment.

Standard English tests can have different layouts and administration formats (computer-based, internet-based or paper-based: PTE Academic, iTEP or IELTS, respectively) and all seek to be internationally recognized and trusted testing systems for both employers and education authorities around the world. The Test of English as a Foreign Language (TOEFL), the International English Language Testing System (IELTS) and the Cambridge ESOL Examinations (Cambridge ESOL) are the most popular English tests globally. However, many others also enjoy a high degree of international recognition, such as Trinity College London Exams (GESE, ISE, SEW), Pearson Tests of English (PTE) and the International Test of English Proficiency (iTEP). Most of these exams are designed to assess different levels of competence, from very basic to highly proficient users of the language, and can even be customised to meet students' specific needs. Thus, Cambridge ESOL have a range of business English examinations, including BULATS (*Business Language Testing Service*) and BEC (*Business English Certificate*); IELTS offers two main types of exams, *Academic* (for university and tertiary education) and *General Training* (for school, work or migration); TOEIC (*Test of English for International Communication*) is an alternative to TOEFL and assesses English proficiency for the global workplace; finally, Trinity College London Exams include several options: *Integrated Skills in English* (ISE), *Spoken English for Work* (SEW), *Skills for Life* (SfL) and *Graded Examinations in Spoken English* (*GESE*).

Most of these tests measure the ability of non-native speakers to understand and use English in real-life settings by examining their competence to understand and produce written and spoken English. Examinees are generally given an overall mark according to their level of performance on the whole range of tasks included in the tests. The majority of tests, which assess the four skills, tend to apply broadly the same weighting to each skill, with the weighting of the spoken component at around 20 per cent or 25 per cent of the overall mark. Thus, no special relevance is given to the speaking skill. In contrast, some language exams have been specifically designed to assess oral skills, such as the Trinity College London

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Graded Examinations in Spoken English (GESE), which measure the candidate's proficiency in speaking and listening, the *Test of Spoken English* (TSE) normally used for employment, graduate assistantships and certification purposes, taking into account the ability of non-native speakers to communicate orally, and the *Spoken English for Work* (SEW), which assesses a candidate's oral skills in a working environment. In these exams, the spoken element constitutes the whole of the final mark.

All these tests tend to be based on achievement through external exams, that is, assessment is not made by the students' teachers but by external examining bodies, such as the Royal Society of Arts or the University of Cambridge Local Examinations Syndicate. In fact, these various exams are generally conducted using similar procedures; students must normally register well in advance, pay a fee (at times quite high), and sit the exam in authorized testing centers on a particular date. Exams may last between 90 minutes and 4 hours, and as regards the marking systems, although the scores may vary widely from test to test, all these English tests are aligned with the six levels of the CEFR (Council of Europe, 2001), with a view to gaining transnational recognition of their language certification. In relation to the period of validity of the certification, it is curious that whereas some are valid for life (Cambridge ESOL), others are accepted for a limited time (two years for IELTS). This is a controversial issue, in that knowledge and command of a foreign language does not remain indefinitely, and learners may fossilize, move backward or backslide, according to the opportunities they have for language practice and reinforcement (Selinker, 1972).

Having given a brief overview of English language tests, we will now discuss the main traits of the spoken modules in the examinations of Cambridge ESOL, TOEFL, IELTS, PTE General, Trinity and iTEP, in order to assess the relevance of the speaking skill in the English qualification industry worldwide. We will look specifically at the spoken modules by examining the assessment criteria employed, the speech-based activities used, and the scoring systems applied to oral skills in these general proficiency tests.

3. SPOKEN MODULES IN GENERAL ENGLISH TESTS: TASKS

In many current English language teaching methodologies communication in its fullest form has become the main aim of the teaching-learning process, and interest in oral tests derives from this growing attention to the listening and speaking skills. For this reason, traditional English language tests have adapted to these demands, and include modules to assess the learners' oral communication skills.

It is commonly believed that "oral tests are qualitatively different from other kinds of tests" (Underhill, 1987: 3) since they include items and activities which are typical and exclusive to them, and are conducted under different conditions from the rest of testing. This section analyses the main speaking tasks used to evaluate the nature and effectiveness of the learners' spoken English.

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Most oral examinations include activities which aim to replicate real-life exchanges, with the level that learners aim to attain determining the type and number of speaking tasks included. Underhill (1987:44ff) deals with a whole list of oral testing techniques and their corresponding variations that can be used in the assessment of spoken language: discussion/conversation, oral report, discussion, role-play, interview, learner-learner description and re-creation, form-filling, question and answer, reading blank dialogue, use of a picture or picture story, giving instructions/ descriptions/ explanation, retelling story or text from aural or written stimulus, reading aloud, translating/ interpreting, sentence completion from aural or written stimulus, sentence correction and transformation, and sentence repetition. Despite such a wide variety of available oral testing techniques, most speaking modules of general proficiency tests follow similar models and include very similar types of tasks. Spoken examinations typically begin with some type of introductory interview (questions about the candidate's life and family), which generally serves as a warm-up and as a means of putting the candidate at ease; the test then moves on to a more complex and independent discussion of a topic (also referred to as a long turn) in which the candidates are required to speak on a particular subject; finally, these tests usually conclude with an interactive task where the candidates must take control and manage a real conversation with an interlocutor. In spite of this general progression in speaking activities (personal interview> topic discussion > interactive task), speaking tasks may adopt different forms to fit the test's purposes and students' specific needs (e.g. academic and professional English tests). Accordingly, the *Trinity College SEW* examinations include telephone tasks in which candidates must lead a conversation over the phone, simulating a specific workplace situation. The table below gathers together the most prominent types of speaking activities used in English proficiency tests examined in the present paper. As can be seen, all these tests follow a similar pattern and do not show remarkable differences:

TYPES OF ORAL TASKS IN ENGLISH TESTS					
CAMBRIDGE ESOL EXAMS	**TOEFL**	**IELTS**	**TRINITY COLLEGE LONDON EXAMS**	**PEARSON TESTS OF ENGLISH GENERAL (formerly London Tests of English)**	**THE INTERNATIONAL TEST OF ENGLISH PROFICIENCY (iTEP)**
-Interview -Individual long turn based on a descriptive task (photo) -Interactive task (pictures) -Topic discussion	-Topic discussion (2 tasks) -Spoken tasks based on short readings, lectures, and conversations (4 tasks)	-Introduction and Interview -Individual Long Turn -Two-way topic discussion	-Introductory Conversation -Topic discussion -Interactive task -Telephone task (SEW)	-Sustained monologue -Topic Discussion -Picture description -Role play	-Short Question-Answer task -Topic discussion (Expressing opinions on an issue of two sides)

Table 1. Types of oral tasks included in general English proficiency tests

Print ISSN: 1578-7044; Online ISSN: 1989-6131

On closer inspection, the oral tasks included in English proficiency tests fall into five main categories: (i) *highly-controlled speaking tasks*, such as reading aloud and short question-answer tasks, (ii) *personal interviews*, (iii) *description tasks*, which may turn into more complex comparing and contrasting tasks, (iv) *topic discussions*, and (v) *role plays*, which may take the form of a telephone-mediated task. Some of these activities are clearly more structured and controlled (reading aloud activities) than others (topic discussions). However, all tasks here focus primarily on oral production itself, rather than on oral spoken interaction. This might be considered a shortcoming of such tasks, in that only part of the spoken language is being evaluated. In fact, the CEFR places an emphasis on this issue by making a clear distinction within the speaking skill between oral production and oral interaction.

We will now deal with the main features of these task categories by providing, where possible, specific examples of each task type, these taken from sample material on the websites of the various tests.

(i) *Structured speaking tasks*, such as short-answer questions and reading aloud activities, are very commonly used in computer-based tests like PTE Academic and iTEP. Reading aloud tasks are highly structured speaking activities designed to assess pronunciation (PTE Academic), intonation, stress and fluent reading. If used in teaching, such reading aloud tasks would most profitably resemble real life situations, that is, reflecting activities in our daily life where reading aloud is most common, including giving instructions and directions, reading notices, warnings, delivering news, etc. In iTEP we find short question-answer tasks designed to show the learners' ability to provide relevant answers to short questions. Candidates hear or read a short topic statement (e.g. *After you complete your studies, what kind of work do you want to do?*) and their task consists in answering coherently.

(ii) *Personal interviews* are normally used as a warm-up activity in which the examiner asks candidates general questions about familiar topics: their hometown, family, work, studies, interests, leisure activities, weekends and future plans. This task is found recurrently in English tests; indeed, exams for IELTS, PTE General,[3] TOEFL, Trinity and Cambridge ESOL all start with an introductory conversation of this type. When talking about the learners' hometown in this personal interview phase, for example, candidates can be asked questions such as *what is the most interesting part of your town/village?, what kind of jobs do the people in your town/village do?, would you say it is a good place to live? Why?* This task allows examiners to assess the candidate's ability to provide information about everyday topics and personal experiences.

(iii) *Description tasks* are used to "assess the ability of the test taker to speak continuously in response to a visual stimulus" (PTE General speaking test guide, 2010:3). The main purpose of these tasks is to elicit specific vocabulary related to the image or images, to assess a learner's ability to describe what they see, and to explain their personal feelings about the pictures provided. They can take the form of an individual task or a pair-work activity

Print ISSN: 1578-7044; Online ISSN: 1989-6131

(also referred to as joint task). The main difference here is that while in individual tasks learners are asked to describe the pictures on their own, in joint tasks two or more candidates are asked to describe their own pictures first and then to discuss the issues illustrated in the pictures together and come to a shared conclusion. A more complex variant of descriptive tasks are *comparing and contrasting tasks* where learners are required to compare and contrast two different photographs or pictures (see Figure 1). In this type of task, candidates are often encouraged to analyse the similarities and differences by using more complex structures. The following task included in the PTE General examinations (level 4) illustrates how examinees are asked to do this:

Now, here are two pictures showing different communities. These pictures are being considered for a book called "Different Communities". Please talk about the pictures and tell me how you think these communities have different lifestyles.

(Hand the pictures to the test taker)

(Allow the test taker to speak for about one minute, then ask this secondary prompt)

Which of these pictures would you choose for the cover of the book called "Different Communities" and why?

(Retrieve the pictures)

Figure 1. Description Task (Source Test: PTE General. Level 4)

(iv) *Topic discussions* are also quite commonly found in English tests. They provide examiners with enough information to assess the learners' ability to speak freely and at a certain length on a given topic, and can be either one-way or two-way discussions. One-way topic discussions are long individual turns in which the candidate is required to talk about a particular issue (e.g. *Describe something you own that is very important to you and say where you got it from, how long you have had it and why it is important to you*). In two-way discussions two or more speakers intervene in the conversation. On some occasions, task cards are supplied with issues to cover, and at times examinees may also be given some time to organise their ideas and prepare their production before they actually state their opinions on

Print ISSN: 1578-7044; Online ISSN: 1989-6131

the topic. The topics vary greatly according to the level of English. One topic of a two-way discussion in CAE is technology and its related issues, including *the advantages and disadvantages of shopping by computer, some people say that computers are helping to create a generation of people without social skills. What is your opinion?, etc*. In general, the more advanced the level, the more abstract the subject areas. At lower levels we find topics such as holidays, shopping, hobbies and sports; at more advanced levels speakers are asked to talk about issues such as rules and regulations, dreams and nightmares, global environmental issues, advertising, lifestyles and the rights of the individual.

Topic discussion tasks give learners autonomy and control over this phase of the examination and allow candidates to show their ability to link sentences together and to structure their speech cogently. In more rigid formats such as computer-based tests, as with iTEP, topic discussions are based on short topic statements that candidates hear or read on a computer screen (e.g. *When a reporter writes a story, sometimes the reporter must interview people confidentially, with the understanding that the reporter will hide their identity. Should reporters be allowed to protect the identities of their sources, or should they be forced to reveal them when the public wants to know? Why?*).

Occasionally, topic discussion tasks may derive from what Luoma (2003) calls *decision tasks*, which involve discussing an issue from different viewpoints and negotiating a final conclusion. They are normally a follow-up to topic discussion activities and are useful as a means of assessing interaction, turn-taking and specific language functions, such as expressing opinions.

(v) *Role plays* are activities in which examinee and interlocutor are given specific roles and a specific situation (Ladousse, 1987). The aim here is to test the examinee's command of certain language functions and linguistic resources. Closely related to this are *telephone-mediated tasks*. They are very commonly found in English tests for special purposes, such as business or work examinations (see, for instance, Trinity College *Spoken English for Work Examinations - SEW*). In such tasks the examinee is given a written prompt with a situation which needs to be addressed (e.g. *You´ve been offered a promotion, but for a number of reasons you feel you must decline. Phone the human resources manager to politely turn down the promotion and justify your decision*); candidates are then required to have a telephone conversation with an official examiner, who is in a different room. It is, therefore, a two-way conversation in a specific situation. Conventional conversational rules, the learners' use of the spoken register, and the way the learner manages the situation are all usually taken into account by the examiner.

As can be gathered from this section, a wide range of tasks exists in oral examinations as a "means [to] elicit a sample of language that can be scored" (Fulcher, 2003: 50). Some are highly controlled, whereas others are of a more flexible nature and allow for the candidates' own managing of the conversation. English tests usually start from more concrete and controlled tasks to give learners enough confidence to speak about their personal experiences,

Print ISSN: 1578-7044; Online ISSN: 1989-6131

before a gradual progression towards free production activities in which learners become more communicative and are responsible for the control of the conversation. It is clear then that from the perspective of assessment, the selection of tasks is crucial, in that the type of activities proposed and the way they are completed by learners determine the evaluation process. This will be considered in the next section.

4. ASSESSING SPEAKING SKILLS

Speaking is one of the most difficult language skills to assess since it must be measured in live interaction (Luoma, 2004: 170).[4] Therefore, examiners should consider a number of pre-defined assessment criteria to help them with this difficult task and allowing them to measure the learners' performances objectively. This is a crucial issue because the selection of these criteria will definitely have a bearing on the validity and reliability of the test (Fulcher & Davidson, 2007: 4-22; Xi, 2008).

If we consider the criteria used for the assessment of spoken language in the different tests studied here, we observe that the candidates' speaking performance is normally assessed on the basis of a set of pre-defined factors: linguistic range, accuracy, speaking delivery, promptness of response and the use of different language functions. These tests can then be described as criterion-referenced tests whose main goal is "to obtain a description of the specific knowledge and skills each student can demonstrate" (Linn & Gronlund, 2000: 43). Table 2 and Table 3 on the next page show the criteria, tasks and marking systems used to assess oral proficiency in the different English tests:

If we look at the assessment criteria more closely, we observe that some tests try to score oral performance on a number of analytic traits (e.g. range, phonological control, etc.) while others consider more general skills (e.g. sustained monologue, turn taking, sociolinguistic appropriateness). However, in spite of the apparent variety of the criteria used for the assessment of spoken language, the differences in the assessment criteria lie mostly in the wording of the criteria themselves rather than in the actual features considered. In fact, the descriptors for spoken language included in the CEFR have been "used as a basis for creating test-specific criteria" (Luoma, 2004: 71). Thus, oral performance is basically measured on the basis of the five main criteria mentioned in the CEFR. The CEFR scale for spoken assessment contains five different analytic criteria: *range*, *accuracy*, *fluency*, *interaction* and *coherence* (Council of Europe, 2001: 28-29).

Print ISSN: 1578-7044; Online ISSN: 1989-6131

EXAM	**CAMBRIDGE ESOL EXAMS**	**TOEFL**	**IELTS**[5]
ORAL PART (WEIGHTING)	20 % to 25 %	25 % 0-30 score scale	25 % 1 -9 score scale
ASSESSMENT CRITERIA	♦ Grammatical Resource ♦ Lexical Resource ♦ Discourse Management ♦ Pronunciation ♦ Interactive Communication **(**Galaczi, 2005: 17)	♦ Speaking delivery ♦ Use of language ♦ Topic development	♦ Fluency and Coherence ♦ Lexical resource ♦ Grammatical range and accuracy ♦ Pronunciation
TASKS	-Interview -Long Turn (photo) -Collaborative task (pictures) -Discussion	-Topic discussions on a familiar topic (2 tasks) -Spoken tasks based on short readings, lectures, conversations (4 tasks)	-Introduction and Interview -Individual Long Turn -Topic discussion
SCORES	Each candidate receives: -A standardized score out of 100 -Summary of performance in each paper on a scale: • Exceptional • Good • Borderline • Weak -A global grade with a recognition of achievement at three levels: A: Certificate (+ level) B: Certificate (=expected) C: Certificate (- lower) D: Fail	Four levels of performance: Good: 26–30 Fair:18–25 Limited: 10–17 Weak: 0–9 Six speaking tasks rated from 0 to 4. Total score: 30	Nine levels of performance: 9: Expert User 8: Very Good User 7: Good User 6: Competent User 5: Modest User 4: Limited User 3: Extremely Limited User 2:Intermittent User 1: Non User 0 : No attempt

Table 2. Cambridge ESOL, TOEFL and IELTS: Main Traits

EXAM	**TRINITY COLLEGE LONDON EXAMS**	**PEARSON TESTS OF ENGLISH (formerly London Tests of English)**	**THE INTERNATIONAL TEST OF ENGLISH PROFICIENCY (iTEP-Plus)**
ORAL PART (WEIGHTING)	50 % to 100 %	25%	20%
ASSESSMENT CRITERIA	♦ Coverage of communicative skills appropriate for the grade ♦ Language functions ♦ Grammatical, lexical and phonological items ♦ Accuracy and appropriacy in language use ♦ Fluency and promptness of response	♦ Fluency ♦ Interaction ♦ Range ♦ Accuracy ♦ Phonological control ♦ Sustained monologue ♦ Thematic development ♦ Sociolinguistic appropriateness ♦ Turn taking (At levels 2–5)	♦ Appropriateness for a particular purpose ♦ Vocabulary ♦ Organization and focus ♦ development ♦ Grammar and mechanics ♦ Pronunciation ♦ Ease ♦ Tone
TASKS	-Introductory Conversation -Topic discussion -Interactive task -Telephone task (SEW)	-Sustained monologue -Topic Discussion -Picture description -Role play	-Spoken response to a short question -Topic discussion (Expressing opinions on an issue of two sides)

Print ISSN: 1578-7044; Online ISSN: 1989-6131

SCORES	Four levels of performance: A: Distinction B: Merit C: Pass D: Fail	Four levels of performance: -Pass with Distinction -Pass with Merit -Pass -Fail Test takers are also provided with a breakdown of their overall score out of 100 and skills scores out of 25 within their Performance Report.	Six levels of performance: 6: Advanced 5: Low Advanced 4: High Intermediate 3: Intermediate 2: Low Intermediate 1: Elementary 0: Beginning Half-levels (2.5, 3.5, etc.) are possible.

Table 3. Trinity College Exams, PTE and iTEP: Main Traits

Range and *accuracy* are measured across all tests examined in this survey. *Range* refers to the lexical repertoire, sentence patterns and formulaic expressions used by speakers in talking about a wide variety of topics. *Accuracy* includes being in control of grammatical structures and using these structures appropriately; and although the CEFR does not mention pronunciation overtly, accuracy also involves knowing how to pronounce words correctly. These two qualitative features have different names in different tests. Hence, *range* is referred to as simply "range" (PTE) or "vocabulary" (iTEP), whereas in some tests it falls under the general category "use of language" (TOEFL), and in others it is split into two different assessment specifications, grammatical and lexical resource (Cambridge ESOL, IELTS and Trinity examinations). As regards *accuracy*, several assessment criteria grids include it as "accuracy and appropriacy in language use" (PTE, iTEP, Trinity exams), others divide it into different discrete factors, including phonological control, tone and grammar and mechanics (iTEP); finally, some refer to accuracy under the heading of pronunciation (Cambridge exams, IELTS).

Another criterion which is assessed in spoken tests is *fluency*, denoting the ability to speak at a normal speed without too much hesitation and produce stretches of language with a natural flow. The tests analysed here include *fluency* as an important criterion for oral assessment. Again, the terms used to refer to this feature vary considerably. Hence, in TOEFL we find the label "speaking delivery" while in PTE "ease" is used together with the explicit label "fluency"; finally, in IELTS we find the item "fluency" together with the issue of "promptness of response".

Interaction is another trait mentioned in the CEFR which is measured in spoken modules of English tests, and refers to the ability to initiate discourse, take turns when appropriate, keep the conversation going, invite others in, and connect one's own contributions naturally to the discourse. However, this feature does not occur as frequently in English test criteria. Cambridge ESOL exams, TOEFL, Pearson and Trinity College allude to it in passing, calling it "interactive communication", "turn taking" or simply "interaction". The absence of this assessment criterion from many other tests confirms once again the relative neglect towards spoken interaction here.

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Finally, *coherence* denotes the ability to use connectors and other cohesive devices to link utterances into clear and logical discourse structure, and it is present in all the tests described in this paper. This feature goes hand in hand with another quality, *cohesion*. These two traits are referred to under different labels, such as "discourse management" (Cambridge exams), "topic development" (TOEFL), "organization and focus" (PTE), "thematic development" and "sociolinguistic appropriateness" (iTEP-Plus).

Thus, we observe that the five analytic descriptors mentioned in the CEFR constitute an invaluable basis for the assessment of oral performance in most English examinations. In spite of the fact that the relevance and importance given to each of these traits varies depending on the nature of the test, the assessment criteria are in essence the same for all tests. In addition, although the fulfillment of each of these criteria is often rated on a task by task basis (analytic rating), speaking scores are normally reported as an overall mark on band scales or score points (holistic rating) that illustrate the learners' overall speaking achievement.

Apart from these analytic features of spoken language, a global oral assessment scale was proposed in the manual to the CEFR published in 2009, which provides suggestions for linking language examinations to the CEFR. This global assessment scale (Council of Europe-Language Policy Division, 2009: 184) contains general assessment statements for each CEFR level. In fact, most test makers have adapted their English tests to the assessment criteria described in this document and in the CEFR generally. Thus, the descriptive guidebooks of these English tests in the main strive to explain the relationship between their scores and the six levels documented in the CEFR. Although the similarities between the assessment criteria in tests and the CEFR are often quite clear, it seems necessary to mention the specific correspondences between the tests under discussion here and the six levels of the CEFR, so as to give a general overview of the diversity of scores in the different exams and the equivalences between them. The following table shows how the different English tests relate to the levels in the CEFR:

CEFR	CAMBRIDGE ESOL	TOEFL IBT	IELTS	PTE (GRAL)	TRINITY (ISE)	iTEP
C2	CPE			LEVEL 5	ISE IV	6
C1	CAE BEC Higher	110-120	8-9	LEVEL 4	ISE III	5-5.5
B2	FCE BEC Vantage	87-109	6.5-7.5	LEVEL 3	ISE II	4-4.5
B1	PET BEC Prelim	57-86	5.5.-6.5	LEVEL 2	ISE I	3.5
A2	KET	32-56	4.5-5.5	LEVEL 1	ISE 0	2.5-3
A1		0-31	0-4			1-2

Table 4. English Tests and CEFR Calibratio

Print ISSN: 1578-7044; Online ISSN: 1989-6131

As can be seen in Table 4, the score bands used for assessing learners' knowledge of English are varied and heterogeneous, and the rating scales used differ considerably across the different tests. Some (TOEFL, IELTS and iTEP) resort to a number of bands to describe levels of competence (e.g. 0-6 or 0-9 band scales) while others (Cambridge ESOL, PTE and Trinity examinations) opt for letters (e.g A, B, C, D). In spite of these scoring differences, the six levels of competence in the CEFR are in fact incorporated in all these examinations so as to "facilitate comparisons between different systems of qualification" (Council of Europe 2001: 21). In relation to the CEFR levels, learners' proficiency in the language is expected to increase as progress through the six CEFR levels (from A1 for total beginners> A2> B1> B2> C1 to C2 for highly proficient students). On the other hand, students' overall scores are based on their command of several analytic features and on their global performance in the test; the final mark depends to a great extent on learners' ability to fulfill the tasks required in the tests. In spite of the usefulness of the descriptors of the six levels of oral proficiency just mentioned, it is true that they may not be transparent and comprehensive enough to be applied in all contexts and situations (Weir, 2005).

5. CONCLUSIONS

This paper has discussed six widely-used English language tests, analysing the main contents of the spoken modules in each of these. We have seen that the spoken skill plays an important role in these general English examinations, in that speaking is weighted equally with other skills, except in those tests which have been specially designed to measure the learners' spoken language, such as SEW.

Oral performance is assessed through different types of tasks, ranging from personal interviews to picture descriptions, topic discussions and role plays, with the aims and the nature of the specific exams determining the activities chosen.

Computer-based examinations show a preference for spoken activities which are quite controlled (short question-answer activities) while face-to-face examinations tend to combine activities of different styles (personal interviews, topic discussions, role plays). In line with computer-based tests and computer-adaptive testing, the use of technology for the assessment of oral production should be carefully considered in the future as it could certainly make an interesting contribution in this area (Chapelle, 2008).

Oral performance is commonly assessed with reference to a number of pre-established criteria, and these can be characterised as analytical traits or as more global abilities. Despite this, test takers are frequently scored on the basis of the five spoken traits mentioned in the CEFR: range, accuracy, fluency, interaction and coherence.

Speaking samples are marked by native English-speaking professionals according to a number of standardised scores and assessment statements which are ultimately linked to the

Print ISSN: 1578-7044; Online ISSN: 1989-6131

six levels set out in the CEFR. Therefore, the analysis of these exams and the tasks included, together with the assessment criteria used, highlight the relevance of the CEFR and its impact on language tests.

In spite of this, our preliminary survey has shown serious weaknesses and limitations in these spoken English tests.

Firstly, it is not always clear which features of the grammar of spoken English are really under consideration; also, oral interaction, which is extremely important in oral communication, seems to be neglected.

Secondly, several documents and guides to these official tests make no mention of mechanisms of updating or adapting to new developments in English language teaching methodologies. Hence, we believe that these tests do not run parallel to the many innovations currently seen in English language teaching; moreover, in some cases they may have a negative wash-back effect on teaching, given that many language schools and other institutions offer specific preparatory courses for these exams (Alderson & Wall, 1993; Bailey, 1996; Cheng, 2008). This question will definitely have to be addressed in the future because of its direct impact on language teaching.

Thirdly, in line with Underhill's (1987) claims, we postulate that within the broad process of designing and establishing evaluation criteria in oral tests, more focus should be placed on those taking tests and those marking them; in other words, oral tests should become more human.

In addition to this, although the CEFR represents a substantial contribution in giving order and coherence to language teaching, in unifying language levels, establishing common criteria and favouring the mutual recognition of diplomas and certificates all over Europe, official language tests should respond more fully to this, especially as regards grading and marking.

Finally, a useful step now would be to expand this exploratory analysis by conducting individual evaluations of the tests reviewed here, as a means of seeing to what extent they meet the requirements of a good language test. The guidelines and checklists provided by Davies (1990) and ALTE might offer a good starting point for this.

NOTES

1. The research reported in this article was funded by the Spanish Ministry for Science and Innovation and the European Regional Development Fund (grant no. FF2012-31450), and by the Galician Ministry of Innovation and Industry and the Autonomous Government of Galicia (CN2011/011 and CN2012/81). We would also like to thank the editors for their useful remarks and suggestions on previous versions of this paper.

2. In this respect it may be useful to consult the following website <http://www.acreditacion.crue.org/mesas.html>, which provides information on the English certificates that are officially recognized by CRUE (*Conferencia de Rectores de las Universidades Españolas*, "Chancellors' Conference of Spanish Universities"). These include certificates

Print ISSN: 1578-7044; Online ISSN: 1989-6131

awarded by the EOI (*Escuelas Oficiales de Idiomas*, "Official Schools of Languages") and those validated by ACLES (*Asociación de Centros de Lenguas en la Enseñanza Superior*, "Association of Tertiary Education Language Centers"). Furthermore, it may also be useful to visit the website of the European Association of Language Testers (ALTE): <http://www.alte.org/>. This association was established by the Universities of Cambridge and Salamanca in 1989, and now has 33 members; it carries out different projects throughout Europe, in addition to having established a set of common standards for its members' exams. Curiously enough, in the case of English, only the Cambridge English Language Assessment and the Trinity College London Exams have passed a rigorous audit and met all 17 of ALTE's quality standards.

3. In Pearson Tests of English, personal interviews are called "sustained monologues".
4. In computer-based examinations, the conditions are slightly different; raters must examine and assess learners' knowledge of the language from the recordings of the speaking samples that learners produce through a headset microphone connected to a computer. This is the case with the PTE Academic test, for example. In such tests, there is no direct interaction between the candidate and the examiner and the tasks are of a more structured nature.
5. The content of the IELTS Speaking test is the same for both version of IELTS, Academic and General Training.

REFERENCES

Alderson, J. C. & Wall, D. (1993). Does Washback Exist? *Applied Linguistics, 14*(2), 115-129.

Bailey, K. M. (1996). Working for Washback: A Review of the Washback Concept in Language Testing. *Language Testing, 13*(3), 257-279.

Chapelle, C. (2008). Utilizing Technology in Language Assessment. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education. Volume 7. Language Testing and Assessment* (pp. 123-134). New York: Springer.

Cheng, L. (2008). Washback, Impact and Consequences. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education. Volume 7. Language Testing and Assessment* (pp. 349-364). New York: Springer.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Council of Europe. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment* (*CEFR*)*: A Manual.* Strasbourg. *< www.coe.int/lang>*

Davies, A. (1990). *Principles of Language Testing.* Oxford: Basil Blackwell.

French, A. (2003). The Development of a Set of Assessment Criteria for Speaking Tests. *Research Notes, 13*, 8-16.

Fulcher, G. (2003). *Testing Second Language Speaking*. London: Longman.

Fulcher, G. & Davidson, F. (2007). *Language Testing and Assessment*. New York: Routledge.

Galaczi, E. (2005). Upper Main Suite Speaking Assessment: Towards an Understanding of Assessment Criteria and Oral Examiner Behaviour. *Research Notes, 20*, 16-19.

Heaton, J. B. (1990). *Classroom Testing*. New York: Longman.

Ladousse, G. P. (1987). *Role Play*. Oxford: Oxford University Press.

Linn, R. & Gronlund, N. (2000). *Measurement and Assessment in Teaching*. (8th ed.). Upper Saddle River, NJ: Prentice Hall

Luoma, S. (2004). *Assessing Speaking.* Cambridge: Cambridge University Press.

Mauranen, A. & Ranta, E. (Eds.). (2009). *English as a Lingua Franca: Studies and Findings*. Newcastle: Cambridge Scholars Publishing.

Selinker, D. (1972). Interlanguage. *International Review of Applied Linguistics in Language Teaching* (*IRAL*), *10*(3), 209-231.

Underhill, N. (1987). *Testing Spoken Language. A Handbook of Oral Testing Techniques*. Cambridge: Cambridge University Press.

Print ISSN: 1578-7044; Online ISSN: 1989-6131

Weir, C. (2010). Limitations of the Common European Framework for Developing Comparable Examinations and Tests. *Language Testing*, *27*, 261-282.

Xi, X. (2008). Methods of Test Validation. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of Language and Education. Volume 7. Language Testing and Assessment* (pp. 177-196). New York: Springer.

ENGLISH LANGUAGE TESTS WEBSITES

- ALTE (Association of Language Testers in Europe):
<http://www.alte.org/> (last accessed October 2013)
- Trinity College London ESOL examinations website:
<http://www.trinitycollege.co.uk/site/?id=263> (last accessed May 2013)
- University of Cambridge IELTS examination website:
<http://www.ielts.org/teachers.aspx> (last accessed May 2013)
- University of Cambridge ESOL examinations website:
<http://www.cambridgeenglish.org/exams-and-qualifications/> (last accessed May 2013)
- British Council IELTS:
<http://takeielts.britishcouncil.org/find-out-about-results/ielts-assessment-criteria>(last accessed May 2013)
http://takeielts.britishcouncil.org/prepare-test/understand-test-format (last accessed May 2013)
- IELTS Home: <http://www.ielts.org/>(last accessed May 2013)
- Pearson English Tests: <http://pearsonpte.com/Pages/Home.aspx> (last accessed May 2013)
- ETS Home for TOEFL: <http://www.ets.org/toefl> (last accessed May 2013)
- Information for iTEP: http://www.itepexam.com/en/itepexams/overview (last accessed May 2013)

Print ISSN: 1578-7044; Online ISSN: 1989-6131