New Challenges for Language Testing

New Challenges for Language Testing:

Towards Mutual Recognition of Qualifications

Edited by

María Luisa Carrió-Pastor

New Challenges for Language Testing:
Towards Mutual Recognition of Qualifications

Edited by María Luisa Carrió-Pastor

This book first published 2016

TABLE OF CONTENTS

INTRODUCTION

MARÍA LUISA CARRIÓ-PASTOR
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

The main focus of this monograph is test development and accreditation requirements and needs. One of the major objectives is to show the key aspects of the application of assessment in higher education and the systems of accreditation. Thanks to its unique perspective, the book offers a different approach to different aspects of foreign language assessment. As universities are one of the best arenas for the analysis of language testing, this book thoroughly prepares higher education teachers to apply pilot studies and shows students' responses to new testing techniques and accreditation requirements. The book offers an enlightening guide for scholars with an academic interest in acquiring the basic principles of language testing and accreditation, with predominantly real cases of how new ways of testing and accreditation can be useful to foreign language teachers and students. Readers will not only come to understand how to use new testing strategies but also have the opportunity to see that the proposals described in each chapter may be useful to language assessment and to the motivation of students as the authors are experienced scholars.

New challenges for Language Testing includes not only reflections on test development but also aspects related to accreditation proposals and the needs and current trends in testing. In this sense, the book is a compilation of chapters that presents the tendencies in assessment and the different ways of developing testing as well as fostering intercultural competence. The reader will discover diverse perspectives of the contribution authors such as the use of assessment and testing as forming part of evaluation. In this book, our main aim is to explore new ways of testing and implementing assessment, but we do not consider that testing and assessment clash. We think testing and assessment are complementary and they should be combined in foreign language evaluation. The teacher or evaluator should decide the importance of testing and assessment when evaluating foreign language proficiency (Canagarajah 2006; Martyniuk 2007; Hornberger and Shohamy 2008). As Fleming (2007:9) explicates about assessment:

In its simplest formulation, assessment provides information on whether teaching/learning has been successful. However the information it provides has a number of potential different audiences whose precise requirements may vary. Classroom teachers need regular information on how pupils' knowledge, skills and understanding are developing, both to inform how they should adjust their teaching and to determine what kind of feedback is needed to improve pupils' learning.

Taking this notion into account, this book sheds new light on the concept of assessment and covers theoretical as well as practical issues. Readers will discover diverse perspectives of the contributing authors on areas such as English for specific purposes exams, assessing writing, listening and speaking skills, linguistic competences, recognition of accreditation peer assessment, e-testing, intercultural competence and local and global accreditation needs. This work will appeal to a wide readership, from language assessors to researchers as well as language centres and universities that promote language testing applied to general and specific foreign language acquisition. The topic of this book is particularly relevant to research in testing skills, accreditation and language assessment in general. The book includes chapters that have been carefully selected and reviewed by a scientific committee, and chosen to offer key recent research on assessment. The double-blind review carried out during the selection of the chapters was important to offer present expert analysis on the central topic of the volume.

This book is divided into two sections. The first section is related to test development and includes five chapters that cover specific aspects of the way assessment can be developed. In the first chapter, Anita Hegedus compares a monolingual and a bilingual English-for-Medical-Purposes speaking test. The author's purpose is to analyse language mediation skills, the introductory conversation, the source language of the input and the assessment of the sub-test.

In the second chapter, Marta Conejero López describes how to assess the speaking skills of business students with B1 level of proficiency following the Common European Framework of Reference for Languages. She focuses on how to assess persuasiveness after training students in persuasive communication. The experiment described by the author emphasizes the role of persuasion and competency acquisition to improve students' effective communication.

In the third chapter, María Boquera Matarredona explains another aspect of linguistic competences by describing the Erasmus+ Online Linguistic Support (OLS). She focuses on the advantages and drawbacks of the OLS as a new opportunity to improve the participants' language

skills. OLS also gives students and trainees the possibility to follow an online language course available in six languages (German, English, French, Spanish, Dutch and Italian) to improve their language competences. The author describes the assessment tool and the different aspects of this initiative.

In the fourth chapter, Elaine Boyd reflects on the formal assessment of academic writing skills. She offers a new approach by presenting ideas from a fellowship scheme in the United Kingdom and using fiction writers to support students. The model explained in the chapter combines assessment with the notions of coherence and storytelling. The author suggests that knowing the voice of students, using peer reading and judgement and connecting stories are potential outcomes for students involved in this model.

In the fifth chapter, María Luisa Carrió-Pastor focuses on peer assessment and its role in foreign language evaluation. She connects this with the importance of motivation when testing the proficiency of students in a foreign language. The results showed that students involved in peer assessment were highly motivated to improve their English and the evaluation they performed was quite similar to the evaluation carried out by instructors.

The second section of the book is also divided into five chapters and focuses on the accreditation requirements and needs that we face nowadays in foreign language testing. In the first chapter of section two, Gillian Mansfield reflects on the use of "mutual" as a key concept in language education as it refers to collaboration as a universal value in Europe. The author presents the way in which the European Confederation of Language Centres in Higher Education works with its members in mutual recognition of their work. She also focuses on the idea of creating mutual awareness and respect of the other in an intercultural Europe.

The second chapter is devoted to accreditation needs. Neus Figueras explains the role of the Common European Framework of Reference in providing objective international standards for languages in Europe. She highlights the importance of testing and assessment associations in facilitating the networking structures to share and exchange expertise. The author pays close attention to the development of new certificates and focuses on what is done in real life and how this affects assessment.

The third chapter explores the state testing system in Russia and the language accreditation model in Spain. Oksana Polyakova and Julia Zabala compare the two models of proficiency exams, contrasting the purpose, test level, preparation course, test construct, test sections, test time, language situations, language skills, rubrics, performance and test

samples. The authors describe the similarities and differences and encourage the mutual recognition of tests.

In the fourth chapter, Aurora Biedma Torrecillas et al. describe the current trends in e-testing, focusing on an online accreditation exam offered by the University of Granada. The test is aligned with the Common European Framework of Reference (CEFR) and complies with international standards of good practice. The authors explain the test, its inspection and administration as well as the validation processes that make it reliable. The *Examen en Línea de Acreditación de Dominio de Español* (eLADE test) was initially created to select participants for mobility programs but nowadays it has become an accreditation test that certifies a CEFR language proficiency level in Spanish.

The last chapter of this volume, written by Cristina Pérez-Guillot and Asunción Jaime Pastor, focuses on B2 listening tasks. The authors explain the importance of proving foreign language competence at Spanish universities. The Language Centre at Universitat Politècnica de València has been closely involved in the development of accreditation exams. In this chapter, the authors focus on the development of listening tasks to certify B2 level following the CEFR. After the analysis and proposal of tasks, the authors conclude that the minimum number of tasks to include in the listening comprehension section should be more than three, and task layout, task order and format were closely related to the candidate's results.

This volume includes ten chapters that will appeal to a wide readership, from those interested in language testing and assessment research to those with an interest in understanding the trends in accreditation and recognition of qualifications. The book is a highly informative and carefully presented volume, providing academic insight for readers with an interest in language testing, assessment and accreditation.

Bibliography

Canagarajah, S. 2006. Changing Communicative Needs, Revised Assessment Objectives: Testing English as an International Language. *Language Assessment Quarterly*, 3 (3), 229–242.

Fleming, M. 2007. The Challenge of Assessment within Language(s) of Education. In W. Martyniuk (Ed.) Evaluation and Assessment within the Domain of Language(s) of Education. Prague: Council of Europe, 9–16.

Hornberger, N. H. and Shohamy, E. 2008. *Encyclopedia of Language and Education*, *Volume 7: Language Testing and Assessment*. Berlin: Springer.

Martyniuk, W. (Ed.) 2007. *Evaluation and Assessment within the Domain of Language(s) of Education.* Intergovernmental Conference Languages of Schooling within a European Framework for Languages of Education: Learning, Teaching, Assessment. Prague: Council of Europe. https://www.coe.int/t/dg4/.../Prague07_Assessment_EN.doc [Accessed 27-04-2016].

SECTION 1

TEST DEVELOPMENT

CHAPTER 1.1

DEVELOPMENT OF BILINGUAL AND MONOLINGUAL ENGLISH-FOR-MEDICAL-PURPOSES EXAMS

ANITA HEGEDUS
UNIVERSITY OF PÉCS, HUNGARY

1. Introduction

The paper aims to compare and contrast a monolingual and a bilingual EMP speaking sub-test conducted at the Department of Languages for Specific Purposes, Medical School, University of Pécs. Since the input for the tasks in the bilingual exam is in Hungarian, the question of whether the paper involves an element of foreign language mediation skills (oral translation) arises. In the case of the monolingual exam, the input for the tasks is provided in English, thus it is inevitable some of the keywords the test taker needs to supply are included, making the exam easier and less reliable for the test taker. Analysis was also directed at the role of the introductory conversation, which, according to a decision made by the Hungarian Accreditation Board for Foreign Language Examination, has to be assessed in the same way as the other tasks of the sub-test. Previously, this part of the exam was not evaluated. The correlation between this task and the other tasks of the monolingual exam was found to be weaker than that between the other tasks.

This paper undertakes to compare and contrast the speaking sub-tests of two English-for-Medical- Purposes (EMP) exams developed by the Medical School, University of Pécs, Hungary: a bilingual and a monolingual exam. Investigations were mainly directed at the role of the introductory conversation, which is the first task of the sub-test, the source language of the input (Hungarian vs. English) and the assessment of the sub-test.

1.1. A bilingual EMP exam: Profex

Profex is an abbreviation for Professional Examination. It won a state accreditation as the first examination for languages for specific purposes in Hungary in 2000. In 2007, the examination levels of Profex were harmonized with levels B1, B2 and C1 of the recommendations of the Council of Europe. The speaking sub-test has three profiles: medicine, pharmacy and general health care. Before 2011, the sub-test included eight profiles (medicine, pharmacy, dentistry, nursing, dietetics, physiotherapy, health education, paramedic science); however, a decision brought by the Hungarian Accreditation Board for Foreign Language Examination reduced the number of profiles of a speaking sub-test to a maximum of three. This decision has resulted in an inevitable loss in the multifariousness of the exam and also in the reduction of the number of test takers practicing in one of the fields dropped from the exam.

The speaking sub-test consists of three tasks: an introductory conversation on the test taker's work and research field, two simulated conversations (one between a health care professional and a patient/client and another one between two health care professionals) and a monologic task, which includes a picture description at level B1, a graph analysis at level B2 and a mini-presentation at level C1. In the simulated conversation between two health care professionals (e.g., discussing a patient's case), test takers are expected to use medical terminology, while in the conversation between a health care professional and a patient/client (e.g., enquiring and examining a patient), they are expected to produce language that also a person unfamiliar with the medical terms could understand. Detailed input is given in Hungarian as it is language that is assessed at the sub-test, not professional knowledge. Background knowledge is a necessary, intrinsic component of the concept of specific purpose language ability (also of English for Medical Purposes) (Douglas, 2000). There is evidence that background knowledge impacts language test performance in tests of languages for specific purposes (Douglas, 2000; Clapham, 1996). The test takers are assumed to possess the professional background knowledge required for taking the exam (otherwise they would not be able to pass) but background knowledge is not scored. Medical professionals are not used in the evaluation of the exam although they are consulted during the test and task development.

1.2. A monolingual EMP exam: Standem

Standem is an abbreviation for Standardized Language Certificate for Medical Purposes. It is a monolingual exam to assess and certify the command of professional English among health-care professionals worldwide. It was developed with the support of the European Commission under the lifelong learning program for the years 2011–2014. A detailed needs analysis and test exams have been carried out but the validation of exam papers is still in progress. The three profiles for the speaking sub-test are medicine, pharmacy and nursing. Task development was based on Profex. The three tasks are: introductory conversation, one simulated conversation between a health-care professional and a patient based on a medical document (e.g., discharge summary) and a graph analysis (levels B1 and B2) or a mini presentation (level C1). As the exam is monolingual, all the prompts are provided in English. Like Profex, professional background knowledge is presupposed but not scored.

2. Assessment of the speaking sub-test

Formerly, the introductory conversation was not scored and served only as a warm-up for the speaking paper. In the Australian Occupational English test, which is also an accredited English-for-Medical-Purposes exam, the introductory conversation is not scored even today (Douglas, 2000). However, a decision by the Hungarian Accreditation Board for Foreign Language Examination in 2011 compels all accredited Hungarian language examinations (even the ones for specific purposes) to assign scores for the introductory conversation. In accordance with this rule, 8 points can be awarded for the introductory conversation, 12 points for the two simulated conversations and 10 for the monologic element (picture/graph/mini-presentation) of the bilingual exam. The scoring is analytical. Since specific purpose language tests are communicative by definition (Douglas, 2000), communicative competence (speech production and comprehension, fluency) forms an important part of the assessment. The other traits along which the scores are assigned are proper use of terminology and grammatical accuracy. They are given equal weight in the evaluation of the simulated conversation, while in the scoring of the other two tasks communicative competence is weighted more heavily.

Since the detailed prompts are given in L1 (Hungarian), the sub-test inevitably involves an element of foreign language mediation skills (oral translation) as the test taker shifts from L1 to L2 and vice versa during the

exam. This element is not manifested in the assessment as test takers are not required to translate the Hungarian prompts into English. They are supposed to use them as clues in a communicative situation.

The traits used in the assessment of the speaking sub-test of the monolingual exam (Standem) are task achievement, discourse/interaction, range/appropriateness and phonetical/lexical/grammatical accuracy. Each is given equal weight and then converted into a scale of 5 in the case of the introductory conversation and into a scale of 10 in the case of the other two tasks.

Standem is monolingual; therefore the prompts are given in L2. This inextricably involves supplying some of the terminology the test taker is supposed to use in L2, thus making the exam easier and less reliable. Giving less input would help to eliminate this problem but would reduce the validity of the sub-task as not providing enough input would entail testing professional knowledge in addition to the language skills the sub-test is intended to assess. However, since at Standem the only simulated conversation is one between a health-care professional and a patient, a crucial element in the assessment is whether the test taker can convey information in a way that a person unfamiliar with medical terminology can understand it. Thus, the test taker is required to transfer information from medical jargon (the language of medical documents) to colloquial language.

3. The introductory conversation

As was referred to above, the introductory conversation, which is the first task of both exams, is assessed in both exams due to a decision by the Hungarian Accreditation Board for Foreign Language Examination. The topics for the introductory conversation (see Appendix) are made publicly available to test takers, therefore they can prepare for this task in advance. Although test takers do not know which topics the examiner will touch upon (two topics have to be included in the case of the bilingual exam and three topics in the monolingual exam), they have the further advantage of initiating the conversation, which provides them with the opportunity of directing the conversation to one of the topics. Being aware of the topics, test takers can prepare sample sentences or a speech on each topic in advance. Since the other tasks of the speaking sub-test are not made publicly available to the test taker and the introductory conversation is the only task that makes preparation in advance possible, the scoring of this task raises validity issues. Therefore, correlation analyses have been carried out to detect the relationship between the introductory conversation

and the other tasks of the speaking sub-test. In the past, correlations have been a useful means to research testing speaking (Fulcher, 1990; Henning, 1987; Meredith, 1990).

3.1. Methods

The sample for the correlation analysis consisted of 100 randomly selected mark sheets (forms on which the scores of the test takers are recorded) of the speaking sub-test at level B2 of the bilingual exam (Profex). Assessment sheets from the monolingual exam were not included in the study because only test exams have been carried out, and thus a sufficient sample of mark sheets was not available.

First, descriptive statistics were done to detect the mean scores and distribution of the results. Subsequently, correlations were calculated between the scores awarded for the individual task and the total score. Finally, correlations were calculated between the scores assigned for the introductory conversation and those assigned for the other tasks of the speaking sub-test (simulated conversation, graph). The statistical analysis was carried out applying SPSS (version 20.0). As the maximum scores assignable for the individual tasks are different (8 points for the introductory conversation, 12 for the simulated conversations and 10 for the picture/graph/presentation), scores were converted to percentages so they were comparable.

3.2. Results

Table 1 displays the descriptive statistics:

Table 1 Descriptive Statistics

	Introduction	Conversation	Graph	Total
N	100	100	100	100
Mean	88.38	67.55	70.27	73.93
Median	100	67	70	77
Mode	100	67	70	87
Degree of skewness	-5.58	-4.07	-2.86	-5.27
Degree of kurtosis	2.3	3.83	0.29	4.09

As is clear from the table, the means of the individual tasks are high, the highest being the introductory conversation. Although the bilingual exam (Profex) is a proficiency test, the department where the exam was developed (Department of Languages for Specific Purposes, Medical

School, University of Pécs) prepares undergraduate medical students for the exam as part of the curriculum and offers preparatory courses to all the test takers (graduates, health workers, researchers). Thus, the exam can be regarded as a special transition between proficiency and achievement tests. In view of this, the means for the conversations and the graph were predictable, but the mean for the introductory conversation falls outside the acceptable range. The degrees of skewness and kurtosis indicate that the data are not normally distributed and a Spearman's correlation should be used rather than a Pearson's correlation.

Tables 2 and 3 summarize the correlation results.

Table 2 Correlations between the Tasks and the Total Score

	Spearman's correlation	Sig. (2-tailed)	N
Introduction - Total	0.743	.000	100
Conversation - Total	0.895	.000	100
Graph - Total	0.910	.000	100

Table 3 Correlations between the Individual Tasks

	Spearman's correlation	Sig. (2-tailed)	N
Introduction - Conversation	0.528	.000	100
Introduction - Graph	0.597	.000	100
Conversation - Graph	0.727	.000	100

There is no agreement in the literature as to when a correlation is considered high or low. According to Dancey and Reidy (2007), correlations between 0.1 and 0.3 are weak, a moderate correlation is between 0.4 and 0.6, and correlations above 0.7 can be considered high. In this case, since all the tasks are intended to measure the same construct (speaking), high correlations would be expected. The correlations between the individual tasks and the total score are acceptable, although the introductory conversation correlates to the total score more weakly than the other tasks. However, table 3 shows that there is only a moderate correlation between the introductory conversation and the other tasks. This questions the validity of the introductory conversation because it is supposed to measure the same construct as the other two tasks, but a moderate correlation suggests that it may not do so. The significance (.000) shows that the findings are not due to chance.

In order to see whether the sample was representative and the results were not influenced by sampling error, a t-test was also carried out. Since the scores were not normally distributed, a non-parametric t-test had to be applied, the results of which can be seen in table 4.

Table 4 Related Samples Wilcoxon Signed Rank Test

Null hypothesis	Sig.	Decision
The median of differences between introduction and total equals 0.	.000	Reject the null hypothesis.
The median of differences between conversation and total equals 0.	.000	Reject the null hypothesis.
The median of differences between graph and total equals 0.	.000	Reject the null hypothesis.
The median of differences between introduction and conversation equals 0.	.000	Reject the null hypothesis.
The median of differences between introduction and graph equals 0.	.000	Reject the null hypothesis.
The median of differences between conversation and graph equals 0.	.025	Reject the null hypothesis.

The table demonstrates that the null hypothesis should be rejected in each case, i.e., it confirms that the results are not due to chance or sampling error.

Effect sizes were also examined to determine the strength of the effect of the correlations between the individual variables. Table 5 illustrates the effect sizes of the result.

Table 5 Effect Sizes

	Effect sizes
Introduction - Total	-0.59
Conversation - Total	-0.5
Graph - Total	-0.3
Introduction - Conversation	-0.57
Introduction - Graph	-0.56
Conversation - Graph	-0.16

Effect size 0.1 indicates a small effect, effect size 0.3 refers to a medium effect, and an effect size of 0.5 demonstrates a large effect (Cohen, 1988). Thus, the correlation between the simulated conversation and the graph has only a small effect, the correlation between the graph

and the total score has a medium effect, and the rest of the correlation pairs have a large effect, i.e., their results need investigating.

3.3. Discussion

The correlation results have demonstrated that the introductory conversation has a low correlation to the other tasks of the speaking sub-test and it has a lower correlation to the total score than the other tasks. This may suggest that this task is not a valid indicator of measuring speaking skills in English for Medical Purposes. However, correlation results have also confirmed that the other two tasks (the simulated conversation and the graph) are valid. A related samples Wilcoxon signed rank test proved that the correlation results are not due to chance or sampling error, i.e., the sample for the investigation was representative. An investigation into the effect sizes of the results revealed that four of the six correlation pairs – including all the correlations of the introductory conversation – have a large effect, which requires investigation and the taking of measures. According to a decision by the Hungarian Accreditation Board for Foreign Language Examination, scores have to be awarded for the introductory conversation, thus this practice will not change despite this study clearly revealing that it impacts negatively on reliability.

One way of overcoming the problem is examiner training. According to Taylor (2011), in a speaking test there is an interaction between the examiner, the test task and the spoken language performance produced by the test taker. Thus, experienced and trained examiners may be able to elicit speech samples that have not been prepared by the test taker in advance. Another solution may be to modify the topics for the introduction and not making them publicly available to test takers. Moreover, the weighting of the traits along which the introduction is scored calls for modification. Scoring posits a significant role in the language testing and assessment context as the third essential component, in addition to a test taker's cognitive abilities and the task/context (Taylor and Galaczi, 2011; Taylor, 2011). In contrast to the scoring of the simulated conversations, where all the traits have equal weight, here twice as many points can be awarded for communicative competence as for the other traits (appropriate use of terminology, grammatical accuracy). Weighting grammatical accuracy more heavily would eliminate some of the reliability problems caused by advance preparation of the test taker.

The same correlation analysis was not carried out for the monolingual exam due to the lack of an adequate sample. Nevertheless, the same

conclusions may be postulated for the monolingual exam as the assessment of the introductory conversation is the same as in the bilingual exam (the topics are available for the test taker and the introductory conversation is scored).

4. Outlook

The speaking sub-tests of two English-for-Medical-Purposes exams have been outlined in this paper. The monolingual exam (Standem) has not yet won state accreditation in Hungary. As the mobility of the workforce is constantly increasing in Hungary, a monolingual language exam is more marketable than a bilingual one. The question arises whether accreditation of the monolingual exam would undermine the bilingual exam. However, for both exams to survive, some of the problems delineated in the paper (scoring of the introductory conversation and the type and amount of input) have to be rethought and adequate measures taken.

Topics for the introductory conversation

Bilingual exam (Profex):

1. Choice of profession and specialisation
2. Research field
3. Topic of Thesis/Dissertation
4. Prevention
5. Daily professional routine
6. Difficulties of the profession today
7. Developments within the profession itself
8. Healthy lifestyle
9. Future plans and prospects for the field
10. Role of foreign languages related to the profession
11. Importance of experience abroad
12. Traditional medicine vs. alternative medicine

Monolingual exam (Standem):

1. Choice of profession
2. Education for health-care professionals
3. Daily routine of the professional
4. Future of the profession
5. Research and development within the profession

6. Health care professionals – patients' relationship
7. Health care system
8. Preventive medicine
9. Role of foreign languages in the health-care profession
10. Conventional medicine vs. alternative medicine

Bibliography

Clapham, C. 1996. *The Development of IELTS: A Study of the Effect of Background Knowledge on Reading Comprehension.* Cambridge: Cambridge University Press.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Dancey, C. P. and J. Reidy. 2007. *Statistics without Maths for Psychology*. Harlow: Prentice Hall.

Douglas, D. 2000. *Assessing Languages for Specific Purposes.* Cambridge: Cambridge University Press.

Fulcher, G. 2003. *Testing Second Language Speaking*. Harlow: Pearson Education Limited.

Henning, G. 1987. *A Guide to Language Testing: Development, Education, Research*. New York: Newbury House.

Meredith, R. A. 1990. "The Oral Proficiency Interview in Real Life: Sharpening the Scale." *Modern Language Journal*, 74 (3), 288–96.

Taylor, L. and E. Galaczi. 2011. "Scoring Validity." In Taylor, L. (ed.) *Examining Speaking: Research and Practice in Assessing Second Language Speaking*. Cambridge: Cambridge University Press.

CHAPTER 1.2

SPEAKING SKILLS TESTING FOR BUSINESS ADMINISTRATION UNDERGRADUATES: HOW TO ASSESS PERSUASIVE SPEECHES IN B1 BUSINESS ENGLISH COURSES

MARTA CONEJERO LÓPEZ
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

1. Introduction

This paper analyses how to test and assess B1 students' speaking skills (SS) in the Universitat Politècnica de València (UPV) Faculty of Business Administration (FADE). Undergraduates learning Business English (BE) in FADE need to be persuasive EFL users in all contexts, particularly in business contexts. They have to succeed when studying B2 English in their 4th year and when using English in presentations, job interviews, etc. In this testing and assessment proposal, the focus is on increasing students' persuasiveness after learning persuasive strategies in class; watching a relevant Polimedia-UPV (videos) reusable learning object (RLO) called "persuasive communication" is fundamental to consolidate persuasion concepts. The author of this paper designed the audiovisual material explaining five strategies to convince (based on Aristotle's rhetoric). The five-strategies learning, testing and assessing process occurs in the first 10 hours of a B1 English for Business course. Two phases were developed: testing and assessing. In the assessment phase, students measure their own persuasiveness, using marking rubrics they had previously designed. Persuasiveness gives EFL users more confidence when facing speaking tests. Undergraduates can find self-assessment of persuasive speeches highly rewarding. Using this research in progress proposal may improve EFL SS testing and assessing processes in higher education centres.

Higher education EFL speaking tests frequently lead to unsatisfactory assessment. FADE undergraduates studying BE need assessment to be rewarding and helpful. Adapting test preparation tasks to transversal competency acquisition could improve BE students' satisfaction with the whole testing and assessing process.

This testing and assessing proposal is research in progress; it focuses on students' persuasiveness training as an innovative way of increasing speaking tests and assessment quality in higher education (HE) courses of BE. After offering a brief literature review, materials and methodology will be discussed. The following sections explain specific procedures to facilitate conducting speaking tests and their assessment with a transversal competency acquisition focus. The last part of this paper will try to predict results for undergraduates participating in the described persuasive speaking and self-assessing experiment, carried out before the end of 2016 in FADE-UPV. Reaching a B2 English level is a must for all UPV students wishing to graduate and HE competency acquisition is crucial for graduates who wish to find a job; training must start in the early stages of each degree. The testing and assessing activities proposed below can bring undergraduates closer to becoming "effective communicators" by the time they graduate through motivating tasks that help them observe, in any given case, the available means of persuasion (Aristotle, 2011).

2. Literature review

Among the relevant references selected when developing this paper, the 2011 edition of *Aristotle's Rhetoric* is included; it is a fundamental piece of literature in any study about persuasive or effective communication. Some of the 21st century researchers who should be examined and cited when analysing assessment processes and Information Technology (IT) for EFL learning and competency measurement are included in the following lines.

Lincoln & Kearney give a solution for the problem of EFL assessment instruments tending to focus on language aptitude. The HE context this paper examines needs specific and sound assessment procedures that go beyond language correctness, as Lincoln & Kearney explain in their 2015 paper; a way of SS testing which matches the peculiarities of FADE undergraduates is required, and measuring persuasiveness is a key objective for reasons that will be examined below.

IT materials are essential in 21st-century universities. For all EFL students in general and for undergraduates studying BE in particular, SS learning and testing tasks are more successful if advantage of IT is taken.

Pérez Guillot & Tudela explain this in an interesting 2012 paper. Considering what Hodgins studied about RLO in 2000, it is helpful to give HE students an online video that can be used and reused (as many times as required) to get specific training; it enhances motivation and creates opportunities for self-study. Blömeke, Gustafsson and Shavelson are also relevant in this research, as well as Jankowska and Zielińska. Their 2015 works about competency measurement in HE and on self-assessment, respectively, have been taken into consideration to develop this paper.

3. Methodology and materials

The methodology and materials chosen for this research are the result of observing the current HE context in terms of student learning needs and aids to satisfy these needs. A look at speech production aids shows that when undergraduates have to produce the script for a speech, they count on spell-checking and grammar-checking software, but when it comes to measuring speech persuasiveness, students have no help; no quick-guides, software or automatic aids exist. Yet, being persuasive and effective is a key transversal competency for all UPV graduates, both in L1 and in EFL (normally, L2). These are the circumstances FADE students have to face; the business context in which they use English intensifies their particular persuasive speech production difficulties. The need to plan a motivating testing and assessing set of procedures is evident; therefore, choosing innovative audiovisual materials, available online, is a way to facilitate undergraduate persuasive communication learning. The resulting competency acquired by students will help them check their own persuasiveness when creating and delivering a speech.

Bearing in mind that "effective communication" HE competency acquisition is a priority for all UPV graduates, persuasive speeches (PS) were chosen in this proposal as relevant speaking test tasks for undergraduates. Specific guidelines in the IT item chosen (Fig. 1) can help students prepare their persuasive speaking test in a motivating way by watching a short video. This new form of working with a RLO, created by the author of this paper, was evaluated as a helpful option (Hodgins, 2000) mainly because, in less than six minutes, the "persuasive communication" online-polimedia/upv RLO teaches students five persuasive strategies based on Aristotle's rhetoric. Each strategy is explained in simple terms; brief examples are provided, plus a link to a well-known Bill Gates speech (Davos, 24 Jan, 2008). Relevant vocabulary and language structures are studied in class before giving students the RLO as homework. Persuasive strategies used by Bill Gates in the first part of his speech will be

identified and commented on; undergraduates will see how he chooses strategies such as questions ("What have I accomplished so far? What do I still want to accomplish?"), contrasts ("economic demand is not the same as economic need") and facts ("Consider that life expectancy has nearly doubled during the last 100 years") to make his speech as persuasive and effective as possible for his Davos 2008 World Economic Forum listeners.

Bill Gates' speech can be understood by FADE undergraduates particularly well as it is an example of persuasive communication in a business context before a business manager and entrepreneur audience; because of their B1 BE level, the students may require intensive work with the rest of this speech at home but no other difficulties are foreseen; in fact, Gates' sentences tend to be short and clear (abundant but straightforward).

After learning the five persuasive strategies for homework and consolidating them in class, undergraduates are likely to start seeing that BE speaking tests may lead to satisfactory assessment.

The approach designed for FADE undergraduates can be summarised as follows:

- To use PS: they are relevant speaking test tasks for FADE undergraduates
- RLO guidelines are useful and motivating to prepare PS
- PS can lead to rewarding assessment: self-assessing persuasiveness
- Self-assessment is appropriate for FADE undergraduates
- Improved HE competency learning will result from following the proposals described in this study

Rewarding assessment is possible if its development engages undergraduates and makes them responsible for measuring their own persuasiveness. Students participating in this experiment will go from assessing BE skills to assessing HE transversal competencies, with "effective communication" being an essential one (UPV requirements link in Appendix B).

The following sections will describe testing and assessing in detail. These two phases, developed in the first 10 hours of a B1 BE course (4.5 credits), are preceded by a UPV-FADE needs analysis during the week before the course starts. Observing specific undergraduate testing and assessing needs is crucial. Two sets of factors have to be examined – B1 BE courses factors and the undergraduates' need for a test that helps them become effective and persuasive communicators, going beyond mere B1 language correctness.

B1 BE courses have testing and assessing specifications in their coursebooks; a high percentage of these are Business English Certificate preparation books for students who may be working and have strictly professional needs (see Appendix C). These students are out of the HE world and their goal is to use English correctly, following the course book recommendations for exam preparation in case they decide to sit the exam; for B1 BE course participants, the Cambridge BEC-preliminary is one of the most appropriate exams. The case of undergraduates is different; their needs are more complex than those of other students and have to be carefully analysed. FADE students are more than BE course participants: they belong to a HE institution and must become effective and persuasive communicators. The Cambridge BEC-preliminary is also an option for HE students, but even if they decide not to sit this exam, passing a B2 English examination will be necessary to graduate.

a. Phase I: Testing

This phase has two steps: pre-testing and testing. The first step is designed to guarantee a successful 3-minute persuasive speech preparation.

- The first 400 words of a 2008 Bill Gates speech is printed and copies are handed out in class before giving students the RLO as homework (Fig. 1). Relevant vocabulary, persuasive strategies and language structures are studied in 1.5 hours of classwork; students finish language analysis at home ("A New Approach to Capitalism in the 21st Century," Davos, 24 Jan, 2008; see Appendix A with the first 400 words of a 4000-word speech).
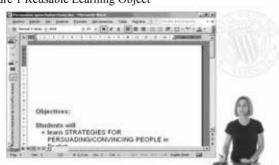
Figure 1 Reusable Learning Object

- Video-watching homework task: "persuasive communication," an RLO (5.45-minute video, available from https://media.upv.es/player/?id=2602e519-4739-9247-9dd7-0fcab3dc57a4& autoplay=true); it can be watched as many times as necessary but three times should be enough. After watching the RLO, undergraduates have adequate training for their persuasive speech preparation and delivery (test).
- Classwork for learning, applying and consolidating five persuasive strategies in 1-minute persuasive speech activities (1.5 hours); only one persuasive strategy is used in each 1-minute speech.
- 3-minute persuasive speech preparation (two hours' classwork; speeches are finished at home if necessary); students try to apply two of the five strategies explained in the RLO.

Testing is the second, and essential, step. Specific tasks are listed in the following lines. Speeches are delivered and recorded in 1st and 2nd practical classes (estimated time for a group of 25 students: one hour of each two-hour practical class). A typed script is submitted and an audio or video-plus-audio file recorded using FADE video-audio lecture-recording equipment (available in most FADE classrooms); students are encouraged to use their own mobile devices in addition to the classroom equipment so as to have an easy-access file to take home.

b. Phase II: Assessing

This phase has a two-hour classwork pre-assessing step. In the first hour, the advantages of self-assessment are explained (Jankowska & Zielińska, 2015) and two existing marking rubrics are analysed. Students select and print existing rubrics; they are free to choose rubrics but they must give reasons to justify their choice; a recommended source is the IELTS website and its speaking exam rubric (Appendix C). The rubric for self-assessment is designed in the second hour (Fig. 2); undergraduates must focus on two questions: "Is my speech persuasive?" and "How persuasive is my speech?"