ESCUELA SUPERIOR POLITECNICA DEL LITORAL

Facultad de Ciencias Sociales y Humanísticas

STANDARDIZATION OF AN ORAL PROGRESSIVE ENGLISH EXAM FOR EFL STUDENTS OF PRE-INTERMEDIATE LEVEL IN AN ECUADORIAN STATE UNIVERSITY: ITS IMPACT AND LEARNERS PERCEPTIONS

PROYECTO DE TITULACIÓN

Previa la obtención del Título de:

MAGISTER EN ENSEÑANZA DE INGLÉS COMO LENGUA EXTRANJERA

Presentado por:

CARLOS DANIEL CAZCO MALDONADO

JUAN CARLOS DELGADO VILLENA

2018

ACKNOWLEDGMENT

First, I thank God, for clearing my mind and my schedule in order to finish this long, overdue research. Blessed be thy name!

Second, all the thanks in the world to my direct and indirect family since your honest and enduring support allowed me to see the real usefulness of ending this master's degree for my own family's future benefit.

Third, I want to give many thanks to my thesis tutor, Jenny. Her speediness while checking the different chapters of this whole document along with her detailed, helpful, professional feedback were and will always be truly appreciated.

Fourth, thank you to my partner in this dissertation, Daniel. His innovative vision brought new and fresh perspectives to this educational project, which after his valuable input became much more than just a required final task to get a fourth level title.

And last, but not least, I need to express all my undefined gratitude to my girls: Denisse, my wife, and Raphaela, my little daughter. They were the real engine thrusting me to the finishing line. All my work and my best efforts will always be for the two of you!

JUAN CARLOS

ACKNOWLEDGMENT

I would like to thank God and the Virgin Mary. They gave me the right time and the correct possibilities to start, continue, and finish the Master Program and our thesis project. They have always been my power, strength and the light of my life that enable me to continue with all my studies and professional projects.

To my whole family and my loved ones, in special to my mom and my dad because they have been the best parents ever and have supported me in all my dreams; to my brother and my little niece, they are my strong commitment for being better every day. To the love of my life who has always been with me during all my thesis project, cheering me up every day since I started.

To ESPOL university, all people from CELEX and their students who gave their best for the success of this project.

To Jenny, the best thesis tutor, ex-boss, colleague, and friend for all her guidance, knowledge shared, support, and help.

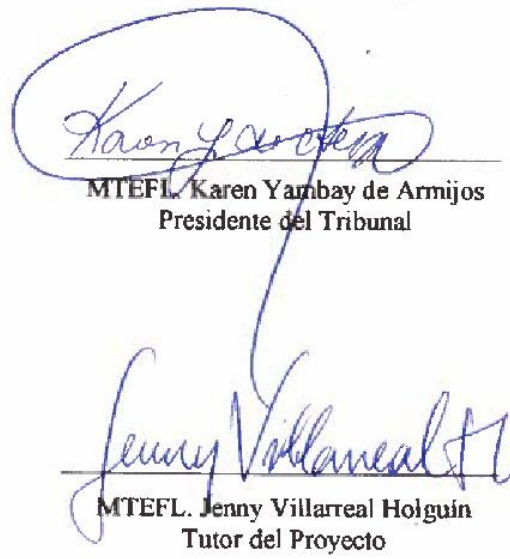
To María Helen Camacho and Karen Yambay, my dear Master Program Coordinators, all their experience and administration permitted me to conclude with this work.

To my friend Juan Carlos, ex-coworker, classmate, and colleague, without his help and willingness to finish our thesis project, it could not have been possible to continue; his vast experience in the field and his professionalism gave the complement to capture our ideas in the project.

To all my friends, classmates, and colleagues from the 7th cohort of the MTEFL Program of ESPOL, the best cohort ever; I learned from their experience and we grew up professionally together.

CARLOS DANIEL

CÓMITE DE EVALUACIÓN



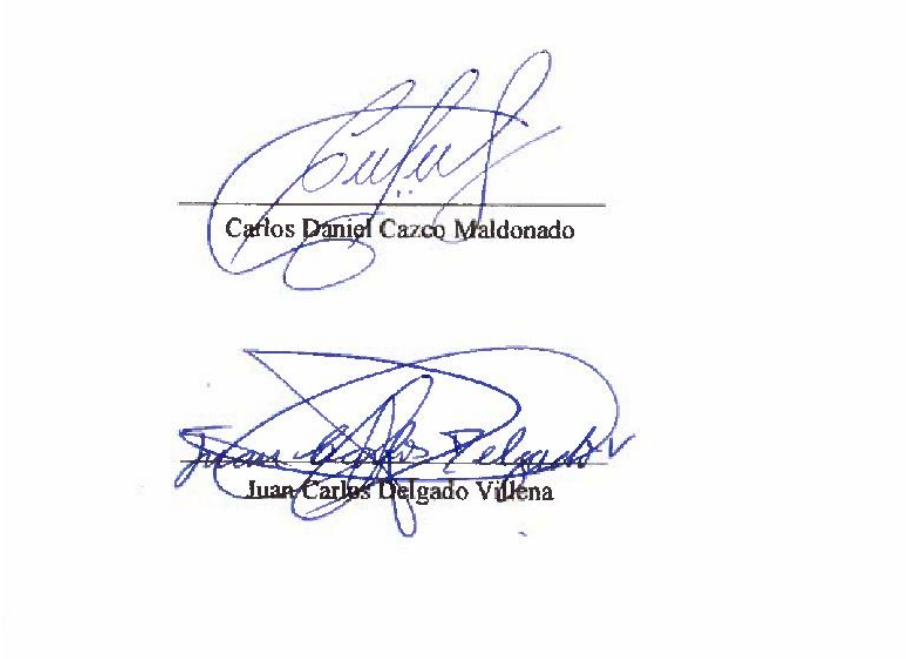
MTEFL. Karen Yambay de Armijos
Presidente del Tribunal

MTEFL. Jenny Villarreal Holguín
Tutor del Proyecto

M.Sc. Jorge Flores
Evaluador

DECLARACIÓN EXPRESA

"La responsabilidad del contenido de este Trabajo de Titulación, corresponde exclusivamente al autor, y al patrimonio intelectual de la misma **ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL"**



INDEX

Abstract

This study describes the tangible impact that the possible standardization of an oral progressive English exam had on the perceptions of sixty-six EFL students enrolled in two courses of the same pre-intermediate level at the language center of an Ecuadorian state university. The academics in charge of this research tested and graded the speaking skill of these learners with the innovative spoken test aforementioned. The qualitative information needed to explore the advantages and disadvantages of this new verbal assessment scheme was gathered when these undergraduates filled out an entry and an exit questionnaire, with twelve questions each, in their L1. Those feedback forms were analyzed by these researchers and their conclusions highlight that their oral exam proposal holds practical and valuable benefits that the current speaking evaluation exercised in this particular education setting does not possess. Additionally, these researchers applied the Student's t-test to the scores these pupils got on their entry and exit vocal examinations for establishing a connection between their actual performances and their opinions about the proposed oral exam. The quantitative data showed that even though many of these students did badly on this new spoken test, they did not state a negative opinion about it. Finally, this educational project should provide the teaching staff of this language center, and any other colleague working in a similar context, with a standardized oral exam that evaluates the speaking skill progression of their students effectively and efficiently.

Keywords: B1 learners, Pre-intermediate students, Critical Thinking, Fluency, Oral Exam.

Resumen

Este estudio describe el impacto tangible que tuvo la posible estandarización de un examen de inglés oral progresivo en las percepciones de sesenta y seis estudiantes de EFL matriculados en dos cursos del mismo nivel pre-intermedio, en el centro de idiomas de una universidad estatal ecuatoriana. Los académicos a cargo de esta investigación probaron y calificaron la habilidad oral de estos estudiantes con la innovadora prueba hablada antes mencionada. La información cualitativa necesaria para explorar las ventajas y desventajas de este nuevo esquema de evaluación verbal se recopiló cuando estos estudiantes completaron un cuestionario de entrada y uno de salida, con doce preguntas cada uno, en su L1. Esos formularios de retroalimentación fueron analizados por estos investigadores y sus conclusiones resaltan que su propuesta de examen oral tiene beneficios prácticos y valiosos que la evaluación oral actual, ejercida en este entorno educativo particular, no posee. Además, estos investigadores aplicaron la prueba t de Student a los puntajes que estos alumnos obtuvieron en sus exámenes orales de entrada y de salida para establecer una conexión entre sus actuaciones reales y sus opiniones sobre el examen oral propuesto. Los datos cuantitativos mostraron que a pesar de que muchos de estos estudiantes tuvieron problemas en esta nueva prueba oral, no expresaron una opinión negativa respecto a la misma. Finalmente, este proyecto educativo debería proporcionar al personal docente de este centro de idiomas, y a cualquier otro colega que trabaje en un contexto similar, un examen oral estandarizado que evalúe la progresión de las habilidades orales de sus alumnos de manera efectiva y eficiente.

Palabras clave: Estudiantes B1, Estudiantes de Pre-intermedio, Pensamiento Crítico, Fluidez, Examen Oral.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1

1. THE PROBLEM

Chapter 1 contains a rapid introduction that is comprised by the number of students that took part in this study, a brief explanation of why this group of learners was intervened, the common traits shared by the EFL instructors involved in this research, a description of the teaching location wherein this research was conducted, and the central questions as well as the sub-research questions around which this study revolved.

This chapter also states the problem by indicating the need this language center had of this study. In addition, it mentions the general purpose and the specific objectives of this research. Likewise, it informs the reader on the significance of the study for this particular Latin American society and educational setting. Finally, it unveils the scope and delimitations of this social discovery.

1.1 Introduction

The following study was carried out for a period of twenty-one non-consecutive days as the dates when the data providers (**N** = 67) were occupied had to be spread out throughout an ordinary class semester of the Center for Foreign Languages from the Ecuadorian state university selected for this research.

This group of Pre-Intermediate EFL (i.e. English as a Foreign Language) learners was intervened with the purpose of determining and describing the impact and perceptions on the experience of having their speaking skill properly and accurately evaluated with an appropriate oral exam that is not only concise and aptly organized, but also promotes dealing with genuine exchange of ideas settings by bringing critical thinking skills into play.

The EFL instructors in charge of this research are both teachers with more than twenty-five years experience in the EFL teaching field combined, who hold international certifications that support their English proficiency levels, who have finished their MTEFL (i.e. Master in Teaching English as a Foreign Language) degree studies, and who had to undergo an exhaustive and rigorous selection process before entering the language center of this state university. The most significant difference

between these social scientists is that one is an experimented oral assessor of international exams while the other is considered an expert in the instruction methodology of "Critical Thinking" by the Ecuadorian EFL teaching community.

To briefly represent the academic setting wherein this study took place, it is imperative to stipulate that the state university where this research was conducted is regarded as one of the most influential public institutions of higher education in Ecuador. Accordingly, it should be no surprise to anyone that its Center for Foreign Languages has been contemplated as a beacon of guidance in the English teaching field for the past twenty-five years. Consequently, with the aspiration of fulfilling the prominent didactic principles advocated by this renowned educational establishment and to withhold its reputation as a reliable leader in the Ecuadorian EFL audience, this language center has been working on the standardization of its grading schemes for the last ten years – incessantly – by stressing the importance of and paying a lot of attention to the productive skills expressions available on all of its "**formative assessments**" and "**achievement tests**" (Roa, 2014) as none of them had been marked within a completely objective framework by its EFL instructors in earlier times.

Prior to closing this short introduction, it is essential to declare the central questions as well as announce the sub-research questions of this scholastic assignment that will not only dictate the behaviour, but justify the decisions of these social scientists during this research.

As they depict what these researchers aspire to grasp, this research project will revolve around these two central questions: 1) what is the impact in EFL students of Pre-Intermediate level at an Ecuadorian state university on having their speaking skill evaluated with a standardized oral exam? 2) What are the perceptions of EFL students of Pre-Intermediate level at an Ecuadorian state university on having their speaking skill evaluated with a standardized oral exam?

There will be three sub-research questions whose importance will be equal as they enclose the secondary points to be examined and explored. For EFL students of Pre-Intermediate level at an Ecuadorian state university, 1) what is the level of approval that using this format of oral exam will have on these learners? 2) What are the main advantages and disadvantages of using this format of oral exam that these learners perceived? 3) What specific recommendations will these learners make on this format of oral exam?

1.2 Statement of the problem

Before this study suggested itself in 2016, the language center selected for it had not yet devised a standardized "summative assessment" (University of Exeter, 2008) for evaluating the verbal dexterity of its students, which is why its personnel had always settled on doing what they deemed as most satisfactory and most pertinent in this regard. In spite of that disconcerting fact, those judgments had continuously been based upon a number of explicit instructions that its prior academic coordinator had put into effect inside this learning setting in 2014. Furthermore, the teaching staff of this language center had already become accustomed to those directives well within the first semester subsequent to her designation for that job post.

Those specific guidelines cited above are: 1) the oral exam should reflect the learners' speaking skills, not memorization. 2) Students should be given between 1 and 2 minutes for in-class preparation before doing the speaking task. 3) Topics must be either found in the textbook or similar to the ones discussed in class. 4) Students cannot do the oral exam on their own as it must be interactive, which means it has to be done in pairs or groups of three or four. 5) The oral exam grade and its feedback must be given to the learners immediately after they have finished the test. 6) The rubrics developed for the oral exam must be applied accurately and meticulously all throughout this summative assessment. (J. Villarreal, personal communication, June 19, 2016)

The foreign language instructors working at this language center have been receiving permanent training for these past four years on the construct of rubrics consequently they are capable of designing and developing rubrics that will work both efficiently and effectively when applied for the grading of a learner's output and functioning in the English language. Unfortunately, this teaching staff has not been offered suitable guidance regarding the handling and conducting of analytic rubrics (Airasian & Russell (2008) as cited in Wadham (2011)) hence it could be affirmed that these professionals are not really ready to use them efficiently and productively since they have not acquired enough experience in these respects, even if they are employed in a state university which counts with and excels at putting these useful and methodical performance appraisal instruments into practice.

However, this shortness of expertise in applying rubrics after ten years of their first introduction in this language center (i.e. 2008) should not be translated into concluding that these workers are unskilled educators as a great part of these EFL

professionals are well-qualified and knowledgeable language instructors – the underlying issue here is that they were never properly introduced to the entire concept of rubrics, especially the analytic kind.

In fact, these researchers are convinced that if these teachers had been properly trained in the application of analytic rubrics, they would have been able to do a much better job when it asked them to grade their students' performance in this foreign language fairly and unequivocally. Hence, finer training for these educators in the use of analytic rubrics is an imminent must as it will be a great endorsement for their upcoming, solid and consistent results while providing marks for the English speaking skills of their students. It is imperative to mention that one of the specific objectives of this research is to provide a more personalized instruction on the proper appliance of analytic rubrics to the EFL instructors of this language center for the apposite assessment of the verbal skill of their students.

The general purpose of this study will be to discover and portray the learners' impact and perceptions on the experience of having one of their productive skills properly and accurately evaluated at the place where they receive their classes and with a suitable summative assessment, which is succinct, follows a specific order and promotes managing real interaction scenarios through the use of analytical thinking.

These next five statements are the specific objectives for this research and they are based on the central or research questions as well as the sub-research questions previously mentioned.

1) To diagnose and establish the students' level of English proficiency in the speaking skill before the research is carried on.

2) To apply guidelines and different techniques to assess oral evaluations in class, and "calibrate" teachers to evaluate according to analytic rubrics.

3) To discern if there is a difference in the development of the speaking skill of undergraduates with B1 (Pre-Intermediate) English proficiency level at an Ecuadorian state university, once the intervention has been concluded.

4) To observe if less skilled learners received better marks in the speaking section of a mid-term achievement test than more skilled learners.

5) To supply the staff of this Ecuadorian state university with data which will serve as proof that using a standardized oral evaluation will certainly change learners' scores.

1.3 Significance of the study

The current law of Higher Education in Ecuador LOES (Ley Orgánica de Educación Superior, 2010), in its Art. 124, requires the mastery of a foreign language for university students; in addition, the RRA (Reglamento de Régimen Académico, 2016), in its Art. 31, demands that all university students must have a B2 proficiency level of English, according to the Common European Framework, to graduate. To enforce those articles in the present educational context of the country, it is required that university graduates know how to use the receptive (i.e. listening and reading) and productive (i.e. speaking and writing) skills that the language has in a manner through which they are able to control these abilities with the final goal of becoming independent users who can interact orally and function fairly well in the professional and academic field. It is also important to mention that this proficiency level is a mandatory requirement to access many national and international postgraduate programs.

The researchers of this project are proposing to help the learners of the language center of this state university to improve their proficiency level of communication, in a particular sense – their speaking skill. Moreover, the results of this study will be a valuable source of information for the EFL teachers of this Ecuadorian state university and for colleagues of other public universities who work in a similar context. Additionally, teachers could have a standardized oral exam to evaluate their students' progression in the speaking skill, following a list of instructions, and determine its impact. Likewise, this research will also show how these students perceived the application of this standardized exam and report on the levels of anxiety they bore during their assessments.

It is significant to mention that having a standardized oral exam will allow students and teachers to be able to understand what the general and specific objectives of this evaluation are and, more importantly, it will help teachers to accurately measure the accomplishment of the learning outcomes proposed by the public university where they work. Furthermore, these Pre-Intermediate EFL students will be more independent and involved in their learning process as they will be continuously engaged on practicing activities taught by their teachers in class.

This verbal appraisal proposal will provide the students of this Ecuadorian state university with a standardized oral exam, which will comply with international and

national standards required by oral assessment processes. This research also seeks to abide the national regulations enforced by the Ecuadorian government as this thesis project is aligned with the "Plan Nacional del Buen Vivir" (2013), specifically with objective 4.8.i, that clearly promotes the learning process of a foreign language within international standards, from early education to higher level education.

1.4 Scope and delimitation

It can be stated that the foremost pursuit of this proposal for a standardized oral exam is to assist the Pre-Intermediate EFL students of this Ecuadorian state university in reaching communicative competence (use) over grammatical competence (usage), as rationalized by Andrews (2000, p.41). The following example should suffice in order to comprehend the real difference between those two terms. For Andrews (2000, p.41), a student who is able to produce a straightforward utterance which has Subject – Verb – Object has proven grammatical competence hence usage. An ENL (i.e. English as a new language) learner (Andrews, 2000, p.xii) focused on use will, on the other hand, engage in gaining the communicative dexterity for conveying thoughts, stating opinions, sharing expectations and expressing desires through the foreign language; in simple words, the ability to *communicate* in that language, effortlessly and fast. (Andrews, 2000, p.40)

These researchers regret reporting that not all of the partakers of this research held the same English proficiency level (i.e. a B1 learner well on its way to B2; B1+, at least) that this language center expects all of its pupils to have for this specific course in relation to CEFR (Common European Framework of Reference for Languages: Learning, teaching, assessment). (University of Cambridge ESOL Examinations, 2011, p.2) Nevertheless, this potential pitfall is what made this collection of various A1 / Breakthrough students, a lot of A2 / Waystage undergraduates (i.e. Basic Users) and some B1 / Threshold learners (i.e. Independent Users) (University of Cambridge ESOL Examinations, 2011, p.4) the best receivers of both the concrete and marginal benefits that participating in this study will allow them to reap.

The conductors of this research anticipate that once all the teaching staff of this language center has truly acquired the essential knowledge required for applying marking standards that are comprehensive, helpful, resourceful and satisfactory (i.e. analytic rubrics), it will help them a great deal while conducting future oral exams.

Furthermore, this pair of social scientists openly endorses the prospective outlook of providing this assembly of EFL instructors with relevant training regarding the use of instructional rubrics (i.e. a concise record of merit stages, devised for grading a fairly dense task, whose main objective is "to give students informative feedback about their works in progress and to give detailed evaluations of their final products" as illustrated by Andrade, 2000, p.13) during class. An ending remark by Andrade (2000, p.13), which is worth mentioning and bearing in mind, regarding this instructive matter: "Instructional rubrics help teachers *teach* as well as evaluate student work. Further, creating rubrics with your students can be powerfully instructive."

It is time to fix the limits of this study hence it is compulsory to provide information regarding the total number of students from the language center of this state university – located in Guayaquil, Ecuador – who participated in this research. The initial sample size was seventy (Original **N** = 70) however, as it is foreseeable, this figure suffered changes due to "experimental attrition" (Lund Research Ltd, 2012). In this particular case, the first number decreased to sixty-seven (Absolute **N** = 67) thus the "experimental mortality" (Lund Research Ltd, 2012) for this research was three (3) participants in total.

Continuing with the marking of boundaries for this research, indicating its length of time is next. This educational project lasted twenty-one (21) days as it began with the first oral exam (mid-term), which took place in the second fortnight of December 2016 (Tuesday 20th), and finished with the second spoken test (end of course) that occurred in the first fortnight of February 2017 (Tuesday 14th). The researchers are only counting the days when they were in direct contact with the learners as it was only then when the participants were intervened.

To conclude with the delimitation of this study, let's move onto the professional bios of the two language instructors (i.e. a Critical Thinking Expert and an Experienced Oral Examiner for International Exams) in charge of this learning revision to become aware of the sort of professionals conducting it. These two EFL language instructors have different curriculums in the English teaching field, but certain points in common: 1) they have been teaching English for more than ten years each, and most of that time, they have worked in institutions of higher education in Ecuador. 2) Apart from being part-time or full-time English language instructors at the scholastic organizations where they have worked, they have overseen administrative posts. 3) They hold international certifications whose scores back up the important fact that their English proficiency

level has always been higher than the one their students have. 4) They have provided other EFL instructors with professionalized training, both within the Ecuadorian national boundaries. 5) They both have official recognitions in EFL teaching from European countries and from the US. 6) They are relatively well-known not only in their corresponding places of work, but in the Ecuadorian EFL community too.

It is quite clear that these two language instructors have a sufficient amount of experience and more than enough qualifications to attempt the attainment of the objectives that this study entails.

CHAPTER 2

2. REVIEW OF RELATED LITERATURE AND PRIMARY STUDY

In order to clarify the rationale behind the different decisions made while preparing and doing this final task, the following section of this dissertation will include the various sources that were observed and analyzed by these language instructors with the objective of gathering their most pertinent, valuable and important information regarding the main concepts and notions that were acknowledged as well as exploited during this research. It will also provide the reader with some key notes concerning the original study, which is an outstanding part in the basis of this educational project.

2.1 Related literature

Introduction

The leading motivation for this oral exam proposal is closing the evaluation cycle utilized for quantifying the English proficiency level (i.e. total standardization of summative assessments that appraise productive and receptive skills alike; in this case, the verbal communication ability) of the learners attending classes at the Center for Foreign Languages of this Ecuadorian state university. It is worth mentioning that the current assessment sequence was first instilled in 2008 and it has been under constant revision by the academic authorities of this language center ever since. Consequently, its permanent development and continuous improvement are common goals for both the authorities aforementioned and the teaching staff who work under their supervision.

This oral exam proposal has two clearly defined sections: 1) an "individual turn" wherein the students have the opportunity to show their actual, distinctive and individual range of fluency (i.e. speaking competence) while doing the task developed for this first part. 2) The second component is an "interactive turn" which comprises an activity that allows the learners to demonstrate their abilities to solve a problem, by either reaching an agreement or recommending other coherent options, and for arguing their viewpoints based on lucid ideas as well as logical suggestions.

This new-fangled spoken assessment scheme categorically requires that the EFL instructors in charge of completing it with the undergraduates of this language center be properly trained on effective, fair, and objective applying of the analytic rubrics

elaborated by this pair of researchers – both immediately and manifestly – ergo those foreign language educators will receive practical, non-theoretical, hands-on teacher training.

The most difficult drawback to handle during this entire research was the negative washback effect, which presented itself during the class sessions assigned for the preparation of this educational project, as it became an inescapable issue for the Experienced Oral Examiner for International Exams. This matter however was not an existing feature in the Critical Thinking Expert coaching sittings.

To make a comprehensible and apparent connection of the four preceding paragraphs with this piece of the dissertation, this literature review has been divided into five main fractions in order to shine enlightening thoughts, provide empirical evidence, and offer solid explanations on the conceptions expressed in those previous lines: 1) why is there a need to have a standardized oral exam at the Center for Foreign Languages of this Ecuadorian state university? 2) What might be the advantages and disadvantages of setting new norms and more objective benchmarks on the spoken tests of the formerly mentioned language center? 3) According to this pair of researchers, concerning this verbal assessment proposal, what are the most essential linguistic topics which have to be mastered in order to "teach" (i.e. develop and practice) speaking as a skill? 4) What has to be worked with these specific set of learners to prepare them for this summative assessment that promotes being orally fluent and managing real interaction scenarios through critical thinking? 5) What areas have to be reinforced in the EFL instructors of this Ecuadorian state university so that they are ready to administer such cumulative evaluation?

1) Why is there a need to have a standardized oral exam at the Center for Foreign Languages of this Ecuadorian state university?

Developing this proposal was the response of an actual and unattended need pending at this Center for Foreign Languages as the speaking skill was the only element of the assessment and evaluation cycle which had not been standardized. The other skills (i.e. listening, reading and writing) were routinely tested and consistently appraised as part of the progress tests (i.e. examinations whose main aim is "seeing how students are getting on with the lessons, and how well they have assimilated what they have been taught over the last week, two weeks or a month" – Harmer, 2007, p.166) as

well as the achievement tests (i.e. assessments which take place at "the end of a term, semester or year" and show "how well students have learnt everything"; they may also be referred to as "exit test" – Harmer, 2007, p.166) that are employed during a normal semester at this Ecuadorian state university.

Characteristics of a good test and their relationship with this oral exam proposal

This pair of researchers are proposing this innovative and effective spoken assessment to replace the existing oral exam format that this Center for Foreign Languages is currently applying to its students as they believe that the present arrangement does not efficiently provide the data it has been designed to, does not convince the learners taking and the EFL instructors marking it that it works properly, and does not "have a positive rather than a negative effect on both students and teachers." (Harmer, 2007, p.167) Consequently, for the people using the existing oral exam format whose feedback is pertinent and compulsory, the present arrangement is not valid (i.e. "it does what it says it will"), does not have face validity (i.e. "when students and teachers see the test, they should think it looks like the real thing"), and its marking scores do not hold reliability (i.e. "anyone marking it should come up with the same results as someone else"), which are central characteristics that a good test should possess as portrayed by Harmer (2007, p.167).

After evaluating the current oral exam format being used at this Center for Foreign Language, another problem that these social scientists identified and tried to solve with this spoken assessment proposal was that even though the existing design might be technically efficient, it is not educationally profitable. According to Wiseman (1961), whose thoughts were advocated by Wall (2000, p.500), the "technical efficiency" of tests ought to not be the only criterion to judge them as adequate or sufficient because educational profitability is also a norm that should be contemplated. Wall (2000, p.500) also resourced to Wiseman (1961, pp. 159-161) for clarifying that 'profitable' tests generate "more 'credits' than 'debits' in the classroom" hence high-quality tests are obliged to comply with the previous effects, not the later ones.

Some of those 'credits' Wiseman (1961, pp.159-161) revealed in Wall (2000, p.500) include that teachers embrace having to deal with the contents of their subjects methodically, follow and complete their courses program of study within an academically fixed schedule, make no coaching differentiation between competent and

ineffective learners, and become aware of standards successfully achieved by other instructors and educational institutions.

In contrast, Wall (2000, p.500) – based on Wiseman (1961, pp.159-161) – indicated that examples of 'debits' incorporate teachers analyzing the backgrounds and personalities of their students to create banks of catered future answers that will have to be memorized, even if they are not understood by the learners, with the intention of using them in exams; instructors being stripped away from their teaching freedom to instill different powers and to tackle established states; teachers completing tasks that have to be done entirely by students because these pupils could not be bothered by those activities; instructors overestimating certain sets of abilities over others since they worked well for improving test performance once; turning teachers into plain marking machines whose exclusive purpose is grading students functioning in that foreign language; and becoming instructors who never consider those contents that will not be appraised in an evaluation, but which are helpful in everyday life.

The oral exam format that these researchers are proposing is in possession of all those aspects maintained by Harmer and it also retains many of those 'credits' shared by Wall within the classroom however it lacks a component that, as Burgess and Head (2005, p.99) explained, is "favoured by exam boards". Although the layout of this verbal communication appraisal makes learners work in pairs (i.e. "**paired testing**"), their performances are not graded by "two examiners present" (Burgess & Head, 2005, p.99) in the classroom; only one instructor does that crucial job given that the institutional context demands it so. As a result, that sole EFL instructor must manage the test and not only work together with the undergraduates constantly (i.e. "role of **interlocutor**"), but take note too on how well each learner is completing the tasks and carrying himself / herself out during those activities (i.e. "role of **assessor**"). (Burgess & Head, 2005, p.99) More details concerning the arrangement of this spoken examination will be better displayed and shared with the reader later on in this chapter.

2) What might be the advantages and disadvantages of setting new norms and more objective benchmarks on the spoken tests of the formerly mentioned language center?

Uses of language tests and their consequences on learners' educational experience

Mostafa and Otham (2007, p.1) agreed with the words of Bachman (2000) to articulate the significance of language tests within any educational context. For this author, "the major uses of language tests are as sources of information for making decisions within the context of educational programs, and as indicators of abilities or attributes that are of interest in research on language, language acquisition, and language teaching." Likewise, Paker (2013, p.1463) has clearly established his firm position regarding this issue: "Assessment is an indispensable part of [the] teaching/learning process." To draw this subject to a close, the reader must be notified on this final remark from Prodromou (1995, p.14) – "tests and examinations–at the right time, in the right proportions–have a valuable contribution to make in assessing learners' proficiency, progress, and achievement."

For Eckstein and Noah (1993), as cited in Wall (2000, p.499), our existing social order calls for tests to satisfy different purposes, such as "allocating sparse places in higher education", "encouraging higher levels of competence and knowledge", and "measuring and improving the effectiveness of teachers and schools", among others. For those reasons abovementioned, tests results can either significantly praise those actors involved in the assessment cycle (i.e. students, teachers, and even schools) with constructive and helpful feedback or condemn them all to negative reviews and comments. (Wall, 2000, p.500) Additionally, Wall (2000, p.500) affirmed that there will be cases when the marks of certain examinations will benefit admissions officers or educational administrators (i.e. decision makers), but they will not be of any gain for the learners and their instructors at all. Finally, if test results in general are purely negative then "anxiety and other negative effects in the classroom" might show up. (Wall, 2000, p.500)

Anxiety and language testing can be easily correlated. As articulated by Pedley (2017, pp.10-11), an experienced language instructor who was the Director of the Center for Foreign Languages where this study took place and whom also carried out a research about Foreign Language Anxiety (FLA) on the students population of this Ecuadorian state university, one of the causes for FLA is "**Language testing** involving particular test formats, or different content and question-types from those practiced (Madsen, Brown & Jones, 1991, as cited in Young, 1991), and tasks and formats that are unfamiliar and ambiguous (Daly, 1991, as cited in Young, 1991)." These researchers

expect this negative effect to decrease considerably once this set of students has received proper training tailored for this well-designed oral exam proposal.

Sayin (2015) has also explored anxiety in speaking exams and how it affects the performance of students. From the results of his research, he observed that

> In language testing, speaking exams may compose the most challenging and stressful part of the testing. Students are tested one-by-one or two as a pair and are expected to talk about on a given task. Thus, students are affected by various factors such as concentration, self-confidence, limited time, and the attitudes of the assessors during the test (Paker & Höl, 2012). (p.113)

Definition of standardized tests

Those two last paragraphs about anxiety and language testing should suffice for now as it is time to start writing about standardized tests. Brown (2004, p.66) stated that it is "important for teachers to understand the educational institutions they are working in, and an integral part of virtually all of those institutions is the use of standardized tests", such comment is definitely right for this Ecuadorian state university and especially true for its Center for Foreign Languages. Consequently, these researchers judge as imperative that the staff of EFL instructors employed in this language center becomes aware of the fact that having detailed knowledge about standardized tests (i.e. "what standardized tests are, what they are not, how to interpret them, and how to put them into balanced perspective in which we strive to accurately assess all learners and their competencies through standardized forms of assessment" – Brown, 2004, p.66) will help them do a better job in general terms when they start using this new format for the oral exam.

Bearing that last observation in mind, this pair of social scientists looked for an explanation that would duly describe what a standardized test is and they agree with the definition provided by Brown (2004) as they find it ample as well as straightforward:

> A **standardized** test presupposes certain standard objectives, or criteria, that are held constant across one form of the test to another. The criteria in large-scale standardized tests are designed to apply to a broad band of competencies that are usually not exclusive to one particular curriculum. A good standardized test is the product of a thorough process of empirical research and development. It dictates standard procedures for administration and scoring. And finally, it is typical of a norm-referenced test, the goal of which is to place test-takers on a

continuum across a range of scores and to differentiate test-takers by their relative ranking. (p.67)

Conceptions in favor of standardized testing

Brown (2004) also inscribed some reasons why he is a frank enthusiast of standardized testing; just three of those explanations will be disclosed now: 1) any standardized test is "a ready-made previously validated product that frees the teacher from having to spend hours creating a test." (p.68) 2) "Administration [of a standardized test] to large groups can be accomplished within reasonable time limits." (p.68) 3) There is an undeniable presence of "an air of face validity to such authoritative-looking instruments." (p.68) Take into account that face validity only happens when "students view the assessment as fair, relevant, and useful for improving learning", according to Gronlund (1998, p.210) as cited by Brown (2004, p.26).

Standardized testing has one of its biggest supporters in Phelps (2008, p.1) since he has not only asserted that it is "the greatest single social contribution of modern psychology", he has also stated that it is "the most useful evaluation method available for human resource intensive endeavors". Even this devotee of standardized tests is aware that they "are not perfect evaluation tools" nevertheless he has declared that, when used "validly and reliably", they have one major, appealing capability in the eyes of decision-makers: "standardized tests provide … useful information that no other evaluation method can provide." (Phelps, 2008, p.1) Furthermore, Phelps (2008, p.2) has appealed to a number of authors (i.e. Gullickson & Ellwein, 1985; Impara & Plake, 1996; Stiggins, Frisbee, & Griswold, 1989; Woodruff & Ziomek, 2004a, 2004b) to prove his statement that with no "standardized tests (or standardized grading protocols) in education, we would increase our reliance on individual teacher grading and testing" however any " teacher's (or school's) grades and test scores are far less likely to be generalizable than any standardized tests' ".

Kelleghan, Madaus and Airisian (2012, p.3) are other adherents of standardized testing and they have asserted that the "superiority of standardized tests over more conventional means of assessment, such as … teacher judgments, in their objectivity and reliability could hardly be doubted". The concept of standardized testing is also backed by Herman and Golan (1990, p.5) for whom such "testing sets meaningful standards to which school districts, schools, teachers, and students can aspire". In addition, the data which those standardized tests provide "can be used as feedback to

shape classroom instruction" (Herman & Golan, 1990, p.5) hence both teachers and students might be benefited from it. Furthermore, this type of testing "makes school systems, schools, and teachers more accountable for student learning." (Herman & Golan, 1990, p.5) Finally, based on the conclusions they reached for a study they did on how standardized testing actually influences teaching and learning, Herman and Golan (1993, p.24) manifested that for "those who agree that the tests represent meaningful learning and instruction, results suggest that standardized testing is a productive policy tool".

Conceptions against standardized testing

No argument is valid or convincing if only one side is heard, which is why the next lines will be devoted to the shortcomings at hand in standardized tests. Frederiksen (1981, p.2) indicated that any "educational test is likely to influence the behavior of students and teachers, provided they know about it in advance" and that is a fact which all exams (i.e. standardized or non-standardized) must deal and cope with. Nonetheless, that factual detail becomes a much more serious deal "if educational tests fail to represent the entire spectrum of knowledge and abilities that ought to be taught" (Frederiksen, 1981, p.2) since that condition establishes "a bias in education, a bias against teaching what the tests do not measure" (Frederiksen, 1981, p.2), and a very likely effect of that preconception is that "the abilities that are most easily and economically measured become the ones that are most taught." (Frederiksen, 1981, p.2) Unfortunately, this trend is happening nowadays because "teachers are being held accountable for what their students learn" (Frederiksen, 1981, p.2) – in most of the occasions – therefore "tests exert pressure to increase time and effort spent in teaching and learning whatever the test measures." (Frederiksen, 1981, p.2)

As a disadvantage on the case of standardized tests, Brown (2004, p.68) mentioned that a problem which might come to light is "the inappropriate use of such tests, for example, using an overall proficiency test as an achievement test simply because of the convenience of the standardization." Another difficulty that was made clear by Brown (2004, p.68) is "the potential misunderstanding of the difference between direct and indirect testing" because a number of "standardized tests include tasks that do not directly specify performance in the target objective", which is an actual limitation of indirect testing. Finally, Brown (2004, p.68) also noticed a grave problem with standardized tests, which is the supposition that these evaluations "correctly assess

all learners equally well. Well-established standardized tests usually demonstrate high correlations between performance on such tests and target objectives, but correlations are not sufficient to demonstrate unequivocally the acquisition of criterion objectives by all test-takers."

Though "standardized testing continues to play a prominent role in educational policy and in efforts to improve the quality of education" (Herman & Golan, 1993, p.20), there are scholars who proclaimed that "the validity and value of traditional standardized tests"(Herman & Golan, 1990, p.6) are plummeting as there is proof which not only shows that "improvements in test score performance [do not] actually signal improvement in learning (Cannell, 1987; Linn, Grave, & Sanders, 1989; Shepard, 1990)" (Herman & Golan, 1990, p.6), but that standardized tests are not completely reliable anymore due to their "narrowness of content, their lack of match with curricula and instruction, and their neglect of higher order thinking skills … (Baker, 1989; Herman, 1989; Shepard, 1990)." (Herman & Golan, 1990, p.6)

Another major disadvantage affirmed by Herman and Golan (1993, p.20) is that instead of having a helpful impact on learners instruction, standardized "testing may trivialize the learning and instructional process, distort curricula, and usurp valuable instructional time (Bracey, 1989; Dorr-Bremme & Herman, 1986; Romberg, Zarinnia, & Williams, 1989; Smith, Edelsky, Draper, Rottenberg, & Cherland, 1989; Stake, 1988)." In addition, educational establishments "serving disadvantaged students are thought to be particularly at risk for such adverse effects (Dorr-Bremme & Herman, 1986)" (Herman & Golan, 1993, p.20), which these researchers regret to report is the present-day distressing reality of this language center for the most part.

Why can't standardized tests be classified only as "good" or "bad"?

As a final point regarding the constructive attributes and the detrimental traits available on standardized tests, Herman and Golan (1993, p.20) reported the results of a study they conducted on "the actual effects of standardized testing on teaching and learning" and they found information that revealed both positive and negative aspects, which merely serves to show that standardized testing has numerous advantages and several disadvantages.

On the latter issue, the preceding authors (Herman & Golan, 1993) put the following claims forward: 1) "Teachers feel pressure to improve student test scores." (p.21) 2) "School administrations give substantial attention to test preparation" (p.21),

which in this case means that the administrators of those schools are constantly on their teachers backs reminding them on the importance of students getting high marks in a given standardized test. 3) "Testing affects instructional planning and delivery." (p.21) 4) "Substantial time is spent preparing students for testing" (p.22) thereby the teaching-learning process suffers because some class time is used entirely for test-preparation activities. 5) "Teachers have doubts about the efficacy (i.e. helpfulness) of testing." (pp.22-23)

On the former respect, Herman and Golan (1993) ascertained these pieces of information: 1) "Nontested subjects also get some attention." (p.22) 2) "Schools give attention to instructional renewal" (p.22) by making efforts to advance their pupils instruction and innovating their teaching practices. 3) "Teachers generally take responsibility for their students' performance" (p.22) because they are aware that their influence has a tangible effect on their learners' accomplishment. 4) Teachers have a posture which is "fairly neutral about the fairness of [standardized] testing."

3) According to this pair of researchers, concerning this verbal assessment proposal, what are the most essential linguistic topics which have to be mastered in order to "teach" (i.e. develop and practice) speaking as a skill?

Before explaining how the speaking skill should be "taught" (i.e. developed and practiced) for this verbal assessment proposal, it is compulsory to define what this ability is and what its core components are. The English language, just like any other modern human language in the world, has four main skills. Two of those are of the productive kind and the other two are of the receptive type; the skill of speaking belongs to the first sort and it "involves using speech to express meanings to other people." (Spratt, Pulverness, & Williams, 2005, p.34)

What speaking components will be considered for this oral exam proposal?

Even though there are some others, which will bear no real importance during this educational project, these social scientists ask the reader to only focus on these three main components present in the speaking skill: 1) interaction, 2) fluency, and 3) accuracy. According to Spratt et al. (2005, p.34), interaction is "two-way communication that involves using language and body language to keep our listener involved in what we are saying and to check that they understand our meaning." The

same authors have defined fluency as "speaking at a normal speed, without hesitation, repetition or self-correction, and with smooth use of connected speech". (Spratt et al., 2005, p.34) As a final point, those scholars have explained accuracy in terms of using "correct forms of grammar, vocabulary and pronunciation." (Spratt et al., 2005, p.34) From these three constituents, this oral exam proposal centers a vast part of its attention on fluency and interaction, which is why accuracy is not seen as the most relevant issue to be evaluated.

"Fluency" defined as "communicative competence" and "speaking effectiveness" for this oral exam proposal

As recently as 2016, two English language teachers from this Center for Foreign Languages carried out a research on the perceptions that their colleagues had on how speaking fluency activities should be planned for their courses and the way those proceedings must be applied in their classrooms. Their conclusions sadly pointed out that many of their colleagues did not have a clear idea on how to develop real communicative competence within their groups as they were not really aware on what the experts on the area have articulated about it. Therefore, this dissertation will include some definitions awarded to this basic term, which is indeed the fundamental item that is graded in the first part of this oral exam proposal.

In 1976, Savignon clearly stated that "communicative competence is not a method" and that it is actually "a way of describing what it is a native speaker knows which enables him to interact with other native speakers." (Leon & Maldonado, 2017, p.27) What is more, as this interaction is both natural and unplanned, it justly asks the interlocutors for a lot more than awareness of the linguistic code. Therefore, a person expected to have seized this linguistic feature of a "native speaker" must not only know "how to say something but what to say and when to say it." (Leon & Maldonado, 2017, p.27)

To draw the opinions of Savignon (1976) to a close regarding this issue and for distinctly defining the concept of communicative competence, this author made especial emphasis on the following two striking points: 1) even though one might be knowledgeable of a linguistic code – certainly "a part of the acquisition of communicative competence as a whole" – this does not mean that one has full command "of language use in personal transactions". 2) For reaching communicative competence, accuracy in using "discrete linguistic elements (i.e. surface features of

language-verb forms, use of prepositions, noun endings,' word order, pronunciation and the like)" is not an essential matter. (Leon & Maldonado, 2017, p.27)

When it comes to the most essential linguistic topics that this oral exam proposal asks students to master, speaking fluency is probably the most difficult item to achieve for weak learners, albeit the presence of accuracy in the same department. Correspondingly, the construction of being orally fluent is what will be better explained now by using definitions related to being communicatively competent in a spoken form. For instance, Susanto (2012) has clearly stated that:

> Understanding the aspects and characteristics of Communicative Competence can help the speaking teachers to guide their learners into speaking atmosphere[s] that make them speak naturally. Negotiation of meaning and management of interaction in communicative competence reflect to the focus on the use of language, not on the usage. Oral communicative tasks given to students are the speaking teachers' consideration to create students' orally natural communication. (p.1)

Despite the fact that "the success of teaching speaking is absolutely emphasized to the use, not the usage" (Susanto, 2012, p.3), an individual who yearns to be communicatively competent must be fully aware of these three areas of knowledge that are deeply intertwined with the skill of speaking: mechanics (i.e. grammar, pronunciation, and vocabulary), functions (i.e. interaction and transaction in addition to interaction/relationship building and transaction/information exchange), and social and cultural rules and norms (i.e. length of pauses between speakers, rate of speech, relative roles of participants, turn-taking). (Susanto, 2012, p.3)

These social scientists decided to cite this whole paragraph written by Susanto (2012) wherein he distinctly explains what communicative competence is because in those accounts the author gives details of its linguistic nature, explains its relationship with language in broad terms, reports on what being communicatively competent is and what it is not, and enlightens the reader on the areas that EFL instructors need to work on so that their learners might achieve communicative competence.

> Communicative competence is a term in linguistics, not only refers to a language user's grammatical knowledge but also social knowledge about how and when to use utterances appropriately. The ability to use the language correctly and appropriately according to communicative competence is to accomplish communication goals. The desired outcome of the use of the language is the

> ability to communicate competently, *not* the ability to use it exactly as a native speaker does. It means that the communicators of the language would communicate naturally without the strict tie of native speaker's influence. This condition really mirrors the existence of communicative competence as the achieved target of learning language. The teachers of language, of course, lead their students based on what is suggested by the communicative competence that involve some areas: linguistics competence, Sociolinguistics competence, Discourse competence, and Strategic competence. (pp.4-5)

Abbaspour (2016, p.146) paraphrases Hymes (1971) to indicate that for speaking effectiveness to take place "L2 learners need to know not only the linguistic knowledge, but also the culturally acceptable ways of interacting with others in different situations and relationships", which is why his theory of communicative competence "consists of the interaction of grammatical, psycholinguistic, sociolinguistic, and probabilistic language components."

To finish, these researchers also acknowledged that a learner trying to become communicatively competent ought to be able to handle himself / herself well during transactional discourse (i.e. "an interaction that focuses on getting something done", according to Richards (2016) in Leon & Maldonado (2017, p.28)) and cope with interactional discourse (i.e. "communication that primarily serves the purpose of social interaction" – Leon & Maldonado (2017, p.28) based on Richards (2016)) more than fairly well.

It is essential to state that in Chapter 4 there is a complete explanation on how oral fluency – speaking competence or communicative competence in this dissertation – was worked with almost half of this cluster of learners (i.e. the students tutored by the Experienced Oral Examiner for International Exams) and why it was done in that fashion. In spite of that fact, this pair of researchers asserts that "**speaking competence**" (i.e. "to enable learners to use English for communication", as portrayed by Abbaspour, 2016, p.144) or "**communicative competence**" (i.e. capability comprising "grammatical competence, discourse competence, sociolinguistic competence, and strategic competence which reflect the use of linguistic system and the functional aspects of communication respectively", like the words of Canale and Swain (1980) advocate in Abbaspour, 2016, p.146) can be determined and well-measured with this standardized oral exam proposal.

These instructors are also keen on emphasizing that attempting to successfully establish and properly quantify degrees of oral fluency inside this collection of students are achievable aims due to the format of this verbal assessment scheme, which is based on an international exam that follows CEFR (The Common European Framework of Reference for Languages) standards (i.e. complying with descriptors whose effects are both effective and long-lasting, as proven by İşisağ & Demirel, 2010, p.191) and fulfills its rigorous criteria for the evaluation of spoken skills since that distinguished test makes effective and efficient use of an assortment of "valid tools for the assessment of oral performance." (Roca, Varela & Palacios, 2013, p.53)

Further ahead, the reader will find out how the linguistic topics of interaction and accuracy were attended and serviced considering the particular format of this oral exam proposal.

4) What has to be worked with these specific set of learners to prepare them for this summative assessment that promotes being orally fluent and managing real interaction scenarios through critical thinking?

Why and how should the skill of speaking be taught in the classroom?

After defining the skill of speaking and its components that were taken into consideration along this dissertation, it is the moment to explain why this productive skill should be taught in class. For Harmer (2007, p.123), there are "three main reasons for getting students to speak in the classroom". Those substantial explanations are: 1) oral pursuits supply students with chances for practicing "real-life speaking in the safety of the classroom", 2) speaking tasks allow learners and their teachers to realize "how successful they are, and also what language problems they are experiencing", and 3) the fact that "students gradually become autonomous language users" due to the constant options they have "to *activate* the various elements of language they have stored in their brains". (Harmer, 2007, p.123)

How can strains of EFL instructors, in any given class, arrive at those key motives put on view by the previous renowned author and mentioned in the paragraph above? According to Harmer (2007, p.132) himself, that process is surely reached once the teacher knows how to stay out of the students' way, without making them feel abandoned during that particular speaking task since learners more often than not certainly "appreciate teacher participation at the appropriate level – in other words, not

too much!" However, how is that done exactly? How does one identify that "not too much!" stage? Harmer (2007, p.132) stated that "there may be times when teachers need to help an activity along through prompting (and perhaps participation), provided it is done sensitively" hence good rapport among learners and the instructor, being sensitive towards the students linguistic setbacks, and being able to empathize with the pupils communicational needs are key factors which endow any teacher with the proper situational thermometer.

What types of tests for assessing speaking were analyzed for this oral exam proposal?

Regarding the skill of speaking, knowing what to teach, why to teach it and how to do it allow EFL instructors worldwide to coach their learners to be prepared for an actual, authentic conversation with another English user or for an illusory, imaginary interaction – quite common in international exams – that has been designed to measure the real and existent English level of a person. Bearing in mind that this research is in itself a proposal for a standardized oral exam including some lines with concepts allocated to the test types applied for assessing speaking is not only reasonable, it is also obligatory. Nonetheless, those utterances will only include ideas about those formats which seem to be ideal for both these researchers and the explicit purpose mentioned above.

Harmer (2007, p.168) maintained that teachers in general can do discrete-item testing (i.e. "testing one thing at a time (e.g. testing a verb or a word)") or resort to integrative testing (i.e. "asking students to use a variety of language and skills to complete a task successfully.") These social scientists decided to go after the second test brand as they created a number of tasks that ask learners to put their English language awareness and command under the spotlight for this novel speaking assessment. Harmer (2007, p.168) also made an unambiguous distinction between direct (i.e. "the one that asks students to do something with language") and indirect test items (i.e. "those which test the students' knowledge of language"). This oral exam proposal makes sure that the students truly display their understanding and comprehension of this foreign language as it exploits direct test items and their core attribute of having "more to do with *activation*" instead of "*study* – that is the construction of language". (Harmer, 2007, p.168)

What are the types of tasks from speaking tests present in this oral exam proposal?

As it could be read above these researchers considered the notions of integrative testing and direct test items for designing this oral exam proposal however there were other considerations related to fitting requirements for speaking tests that were also reflected and drew upon for this educational project, such as tasks types.

According to Burgess and Head (2005, p.99), speaking exams contain "a series of short tasks" and each one of those tasks has been "designed to demonstrate a different function of the spoken language." Burgess and Head (2005, pp.99-103) also named the four task types upon which the majority of oral evaluations are based as "**interview tasks**" (i.e. tasks that test the candidates ability "to provide general personal information"; also known as "'question and answer' task"), "**presentation tasks**" (i.e. a task that makes candidates speak "at length, usually for between one and three minutes, on a prescribed topic" and such oral production must be done "in an appropriately fluent and coherent manner"), "**negotiation tasks**" (i.e. tasks "in which candidates discuss a given situation in order to reach a decision"; learners "may be required to make suggestions, discuss alternatives, find differences, put items in order, or speculate about a situation" for these tasks), and "**discussion tasks**" (i.e. a task which relies on "one or more open-ended questions to generate discussion with the candidate(s)" and gives "an opportunity for the interlocutor to intervene directly"; additionally, its "discussion topic is likely to link thematically with the earlier activities").

Before deciding on the final format for this spoken assessment scheme, these teachers had a debate on whether the second task should be a discussion task instead of a negotiation task. Even though the former provides pupils with "a final opportunity to show themselves and their speaking abilities at their best" (Burgess & Head, 2005, p.103) by expressing their points of view and conveying their standpoints, these researchers agreed on the fact that the latter would be more challenging for the learners – due to their English proficiency level – and much more interesting to grade for these EFL instructors.

Definitions of critical thinking and the ideal critical thinker

In order for students to have the essential linguistic tools to do quite well in the second part of this oral exam proposal wherein negotiation tasks will be taken advantage of and problem-solving expertise development hauled along with them, these

researchers made the decision of introducing the concept of critical thinking and its skills to these pupils.

Many authors have delineated what critical thinking is (i.e. Ennis, 1985; Facione, 1990; Allen, 2014) and what its practical role in our lives is (i.e. Paul, 1992; Carroll, 2007; Lai, 2011). For Allen (2014, p.37), this singular concept can be specified as "a habit of mind characterized by the comprehensive exploration of issues, ideas, artifacts, and events before accepting or formulating an opinion or conclusion." In contrast, Facione (1990, p.6) resourced to a panel of experts to provide a concise description of what Critical Thinking is, and they indicated that CT is "purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based." Similarly, the comment Ennis (1985, p.45) made to define Critical Thinking sustains that it is "reflective and reasonable thinking that is focused on deciding what to believe or do"; he also maintains that unlike the skills of higher-order thinking, critical thinking holds a much more evident concept. (Ennis, 1985, p.45)

The assertions of Paul (1992, p.8), in Reid (1998, p.18), establish that critical thinking, "in contrast to rote memorization or simple information recall, has as its goal, the simulation of analytical and evaluative processes of the mind". Lai (2011, p.2) has also emphasized that critical thinking greatly stimulates our intellects by stating that CT contains "the component skills of analyzing arguments, making inferences using inductive or deductive reasoning, judging or evaluating, and making decisions or solving problems"; in fact, for this author, critical thinking engages "both cognitive skills and dispositions." Equally, Carroll (2007, p.5) shone light upon the thoughts of Richard Paul and others too with the intention of designating critical thinking like "thinking that is clear, accurate, knowledgeable, reflective, and fair" for determining what to accept as true or what action to complete. When the conceptions of these three authors are combined, the end result is the description of "the ideal critical thinker" who Ennis (2011, p.5) has portrayed as an individual that "has the ability to clarify, to seek and judge well the basis for a view, to infer wisely from the basis, to imaginatively suppose and integrate, and to do these things with dispatch, sensitivity, and rhetorical skill." Summing up, critical thinking helps us to make better decisions because our thinking process becomes more organized, more logical, and more analytical thanks to the development of its skills.

Critical thinking skills and their development in an ordinary classroom

According to Ennis (2011), there are fifteen critical thinking abilities in total, which will now be presented to the reader:

> 1. **Focus** on a question. 2. Analyze **arguments**. 3. Ask and answer clarification and/or challenge **questions**. 4. Judge the **credibility** of a source. 5. Observe, and judge **observation** reports. 6. Deduce and judge **deduction**. 7. **Make material inferences** to **generalizations** and to **explanatory hypotheses**. 8. Make and judge **value judgments**. 9. Define terms and judge **definitions** using appropriate criteria. 10. Attribute **unstated assumptions**. 11. Consider and reason from premises, reasons, assumptions, positions and other propositions with which they disagree or about which they are in doubt, without letting the disagreement or doubt interfere with their thinking (**"suppositional thinking"**). 12. **Integrate** the dispositions and other abilities in making and defending a decision. 13. Proceed in an orderly manner appropriate to the situation: a. Follow **problem solving** steps. b. Monitor their own thinking (that is, engage in **metacognition**). c. Employ a reasonable critical thinking **checklist**. 14. Be **sensitive to** the feelings, level of knowledge, and degree of sophistication of **others**. 15. Employ appropriate **rhetorical strategies** in discussion and presentation (oral and written), including employing and reacting to "**fallacy**" labels in an appropriate manner. Examples of fallacy labels are "circularity," "bandwagon," "post hoc," "equivocation," "non sequitur," and "straw person". (pp.2-4)

These researchers deem imperative to stipulate that, in line with the notions provided by Ennis (2011, p.2), all those skills can be divided into six categories: basic clarification, the bases for a decision, inference, advanced clarification, supposition and integration, and auxiliary abilities. The first three (i.e. numbers 1, 2 and 3) skills entail "basic clarification", the next two (i.e. numbers 4 and 5) support "the bases for a decision", the subsequent three (i.e. numbers 6, 7 and 8) exemplify and reflect "inference", numbers 9 and 10 elucidate "advanced clarification", numbers 11 and 12 assist for "supposition and integration", and the last three (i.e. numbers 13, 14 and 15) are "auxiliary abilities".

However, not all of those abilities can be effectively provoked or efficiently enhanced within a classroom due to time constraints. Therefore, Reid (1998, p.18) invoked those proclamations of Norris and Ennis (1989, p. 14) to provide a list of those critical thinking skills that can be productively developed and advantageously practiced

in any classroom: elementary clarification (i.e. basic clarification), basic support (i.e. the bases for a decision), inference, advanced clarification, and strategies and tactics (i.e. auxiliary abilities).

Based upon what was previously stated about critical thinking skills and their development in an ordinary classroom (i.e. fundamentals taken from Reid, 1998, p.18), these EFL instructors were able to recognize that to expand the "elementary clarification" skill in learners, they should do activities which make them focus on questions, analyze arguments and ask and answer questions that clarify and challenge. For increasing the "basic support" skill, students have to do tasks wherein they judge the credibility of a source, and make and judge observations. The skill of "inference" will be boosted with exercises that ask pupils to make and judge deductions, inductions and value judgments. In order to enhance the "advanced clarification" skill, teachers must introduce class activities that require learners to define terms and judge definitions as well as to identify assumptions. Finally, tasks whose main aim is to decide on an action, and exercises that promote interacting with others, acutely assist in enlarging the skill referred to as "strategies and tactics".

How can any teacher become an "ideal critical thinking instructor"?

To further the most correct development of those skills that can be toiled in a classroom, these EFL instructors had to become the embodiment of the "ideal CT instructor" (Facione, 1990, p.36); such evolution, as it can be easily foretold, requires doing many activities. Facione (1990) mentioned all of them, but these researchers only focused their attention in a few, and they were: To…

> … help students elaborate, transfer and generalize these skills to a variety of contexts.
>
> … create a classroom … environment which is supportive of CT.
>
> … model CT in [their] teaching ... .
>
> … provide [their] students with thought provoking subjects to learn about, and projects to undertake.
>
> ... evaluate each student's progress, achievement or proficiency in CT continuously. (p.36)

How can critical thinking skills be graded with this oral exam proposal?

As CT is the harnessed tool which will help this set of students to perform well in the most communicative portion of this innovative spoken test, these researchers were obliged to understand how CT is correctly appraised. Critical Thinking is properly assessed when the teacher is aware that s/he "must not simply reward [students] arriving at correct answers" (Facione, 1990, p.33), s/he has to analyze how (i.e. "by way of good CT" – Facione, 1990, p.33; in other words, by looking for models that support ideal critical thinker behavior) they engendered those responses. Finally, when one evaluates the acceptability of a CT assessment instrument, "one should consider content validity, construct validity, reliability, and fairness" (Facione, 1990, p.33), just like the creators of this oral exam proposal did while drawing it.

To finish this thorough explanation of why these researchers decided to bring critical thinking into play for this innovative verbal assessment, it is crucial to proclaim that these EFL instructors expect substantial improvement in interaction and enhancement in accuracy too, but to a lesser degree, due to this particular brand of linguistic training.

How was the speaking component of "interaction" measured in this oral exam proposal?

Referring back to the format that was indoctrinated for this oral exam proposal, it encloses a combination of presentation tasks and negotiation tasks. For the "individual turn", the learners are given a "**verbal prompt**" (i.e. a topic and perhaps … a list of points to help the candidate to focus on suitable content and structure – Burgess & Head, 2005, p.101) that is occupied by the teacher carrying out the "**interlocutor**" role (i.e. "someone who participates in a conversation with a student" – Harmer, 2007, p.167) in this activity. Then again, the format provides the students with a "**visual prompt**" (i.e. a picture or set of pictures, a chart, or diagram – Burgess & Head, 2005, p.100) that students must bring into play during the "interactive turn". These two traits pertain to presentation tasks and, in this case, they are used so that pupils show they can do a task which "demands different skills from those needed for simply asking and answering questions." (Burgess & Head, 2005, p.101) Likewise, these EFL instructors also capitalize on the positively inherent collaborative quality (i.e. these activities command "students to discuss the [specified] situation with each other" – Burgess &

Head, 2005, p.102) bestowed upon negotiation tasks for the "interactive turn" of this oral exam proposal.

Due to their effective layout, it is the deep belief of these researchers that apart from all the training these learners received to implement critical thinking skills while performing in the second half (i.e. negotiation task where problem-solving knowledge is exercised) of this novel speaking assessment, the linguistic issue of interaction was thoroughly and efficiently taken care of when this particular array of students was compelled to get through those specific tasks aforementioned (i.e. presentations, negotiation activities and discussions) in class.

Subsequent to considering the educational setting wherein this standardized oral exam proposal will be used, these social scientists concurred that neither TOEFL iBT (Test of English as a Foreign Language Internet-Based Test) nor TRINITY's GESE (Trinity College London Graded Examinations in Spoken English) have formats which are suitable for measuring this conversational concern well as the scored interaction occurs only with the examiner or language instructor in charge of grading the performance of the candidate. Therefore, there is no real-life-like interaction with another interlocutor whose English proficiency level is similar to the one the candidates have to demonstrate. Nevertheless, the communication that this standardized oral exam proposal fosters has been proven to be generated by the format employed by CAMBRIDGE's PET (Cambridge English: Preliminary qualification {Preliminary English Test – **PET**}) as it promotes interaction among candidates, not a candidate-examiner interface consequently that international, B1, low intermediate exam is the basis for this new verbal assessment.

How was the speaking component of "accuracy" facilitated to the learners for this oral exam proposal?

Accuracy was the last linguistic topic to be mentioned while declaring which main components of the speaking skill were going to be considered during this educational project. These researchers regret to report that they could not focus too much on accuracy because of the actual English proficiency level these students hold, which was not up to the standards of their current course. However, all the presentation tasks, negotiation exercises and discussion activities that were done in class to prepare these learners for this new spoken assessment had feedback phases wherein grammar correction was applicable, not vital though. Besides, according to Burgess and Head

(2005, p.121), the "grammar, vocabulary, topic areas, and language functions tested [in low-level exams, such as this new verbal appraisal] are practised in general elementary and pre-intermediate coursebooks, so little specific exam preparation is needed." Even so, while these learners were doing this oral exam, many of them (mostly from the collection of the Critical Thinking Expert; very few from the assembly under the supervision of the Experienced Oral Examiner for International Exams) proceeded in an systematic way that was suitable to the circumstance, not only by applying problem-solving tactics, but by considering the beliefs and the comprehension stage of their classmates too. These students also made use of idioms or language chunks to be noticed in the presentation activities (i.e. a funny remark) and make a statement in the negotiation tasks (i.e. a relevant and pertinent question). These last points are the same as being "**sensitive to** the feelings, level of knowledge, and degree of sophistication of **others**" and employing "appropriate **rhetorical strategies** in discussion and presentation (oral and written)" (Ennis, 2011, p.4) therefore it can be rightly uphold that accuracy was also enriched as a fringe benefit, not a main aim.

Recapitulating some of its main ideas and to conclude this fraction of the literature review, the format of this oral exam proposal relies on the "individual turn" for demonstrating the speaking competence or oral fluency level of these learners, and to corroborate interaction examples wherein critical thinking skills are channeled, the design resorts to the "interactive turn". Samples of accuracy, on the other hand, are bound to happen throughout both turns.

5) What areas have to be reinforced in the EFL instructors of this Ecuadorian state university so that they are ready to administer such cumulative evaluation?

Definition of rubrics, their main characteristics, and their core components

According to Harmer (2007, pp.171-172), subjectivity certainly is a serious yet foreseeable difficulty that will show up while using direct items in a test, but it might be overcome once that opinionated quality is amended for the performance appraisal process. These researchers have epitomized that correction through the development and application of a marking scheme, which is anchored in specific criteria, that assesses and values the various ranges in the performance of students. This grading tool which makes the evaluation process much more objective and uses "marking scales for

a range of different items" (Harmer, 2007, p.172) is also known as rubrics. Actually, rubrics are being used in higher education to improve student accomplishment, for better teaching and to assess curricula. (Reddy & Andrade, 2010, p.444) Consequently and in order to use rubrics well, the person working on and with them must be knowledgeable about this useful assessment tool and also trained strictly on how to design, apply and improve it frequently and suitably.

Let's start this fraction of this literature review by conceptualizing what rubrics are. Wadham (2011, p.1) has stated that rubrics are "documents that describe varying levels of performance from excellent to poor and show where on that scale a student is achieving a particular learning standards, goal, or objective." Mianto (2012, pp.2-3) refers to Arter and Chappuis (2006) as well as Asmus (1999) in order to corroborate what rubrics are. For the first two authors, "rubrics identify complex, meaningful tasks and allow for consistent judgments regarding the quality of student work." The third author, in contrast, points out that "rubrics are a guideline for rating students' performance." Finally, Allen (2014, p.1) designates rubrics as the provider of "the criteria for assessing students' work", which are usually "used to assess virtually any product or behavior, such as essays, … oral presentations, … and group activities."

At the moment, the reader knows what rubrics are therefore educated thoughts on their educational profitability must be shared immediately. Allen (2014, p.1) does not only specify that rubrics should be employed "to clarify expectations to students, to provide formative feedback to students, [and] to grade students …", she makes especial emphasis on the fact that any "rubric communicates what the outcome really means because it specifies the criteria for assessing its mastery." In addition, according to Petkov and Petkova (2006) and Reitmeier, Svendsen, and Vrchota (2004), as paraphrased in Reddy and Andrade (2010, pp.444-445), "involving students in the development and use of rubrics or sharing an instructor-developed rubric prior to the submission of an assignment was associated with improvements in academic performance".

Now, the reader will be enlightened on the parts that are comprised in a rubric. These EFL instructors estimate that the next explanation, provided by Reddy and Andrade (2010), is the most accurate and professionalizing on the subject of what the parts of a rubric are and their corresponding functions:

> A rubric has three essential features: evaluation criteria, quality definitions and a scoring strategy (Popham 1997). Evaluation criteria are the factors that an

> assessor considers when determining the quality of a student's work. Also described as a set of indicators or a list of guidelines, the criteria reflect the processes and content judged to be important (Parke 2001). Quality definitions provide a detailed explanation of what a student must do to demonstrate a skill, proficiency or criterion in order to attain a particular level of achievement, for example poor, fair, good or excellent. The quality definitions address the need to distinguish between good and poor responses, both for scoring purposes and to provide feedback to students. Scoring strategies for rubrics involve the use of a scale for interpreting judgments of a product or process. (pp.435-437)

Types of rubrics

For the reader to have a clear picture of them, naming the different types of existing rubrics is germane. According to Mianto (2012, pp.5-6), based on Arter (2000), there are four types of rubrics, which are used to grade the performance of learners, and they are holistic rubrics (i.e. rubrics that "provide a single score based on overall impression of a student's performance on a task"), analytic rubrics (i.e. rubrics which "provide specific feedback along several dimensions" therefore "the teacher is able to get more detailed feedback"), general rubrics (i.e. rubrics that "contain criteria that are general across tasks"), and task specific rubrics (i.e. rubrics which "are unique to a specific task"; its main advantage: "more reliable assessment of performance on the task.")

Why should rubrics be implemented to grade the speaking skill?

At this point in time, the reader will come across a list, deducted by Mianto (2012), which illustrates the reasons why rubrics should be implemented to grade the speaking skill:

> Rubrics help to improve student performance and increase learning of speaking skill.
>
> Using rubrics in speaking will reduce grading time.
>
> Rubrics can be a proof to be shown to [whoever is interested] related to the students' work.
>
> When the criteria for grading a speaking assessment are descriptive and explicit, it helps to reflect the weighted importance of the speaking assignment's objectives. (pp.8-9)

Validity and reliability of rubrics

Everything related to general knowledge apropos rubrics has been clearly defined, but how can one know that rubrics are valid and reliable tools for grading the performance of students? Falling back on Reddy and Andrade (2010, p.441) who support their conjectures in Moskal and Leydens (2000) as well as Fleenor, Fleenor, and Grossnickle (1996), the *"reliability of rubrics"* can be determined by the existence or absence of rater reliability. Therefore, inter-rater reliability (i.e. "the consistency of scores that are assigned by two independent raters") as well as intra-rater reliability (i.e. "the consistency of scores by the same rater at different points in time") have to be proven extant. However, these two conditions are intrinsically intertwined with two specific requirements, which are consensus (i.e. prerequisite which "measures if raters assign the same score"; also called "agreement") and consistency (i.e. stipulation that "provides a measure of correlation between the scores of raters") therefore it is more important to attest the subsistence of these two specifications. On the contrary, the *"validity of rubrics"* is closely related to the clarity of the language it contains and manages since "an ambiguous rubric cannot be accurately or consistently interpreted by instructors, students or scorers (Payne 2003)" as exposed by Reddy and Andrade (2010, p.443).

Although validity and reliability serve the same purpose, one is more important than the other. If there is inter-rater reliability after maneuvering a certain rubric, then "a relatively common interpretation of student performance" (Reddy & Andrade, p.445 based upon Simon and Forgette-Giroux 2001; Hafner and Hafner 2003; Dunbar, Brooks, and Kubicka-Miller 2006) can be assertively declared thus outstandingly effectual rater training is vital for a good application of rubrics.

How was the whole construct of rubrics utilized during this educational project?

Once all the needed information regarding the subject matter of rubrics has been imparted to the reader, these researchers will start indicating how these valuable assessment tools were effectively and efficiently applied on this oral exam proposal.

First of all, this innovative verbal evaluation recognizes that "the teacher is the most important agent of assessment" (NCTE – National Council of Teachers of English, 2010) since this EFL instructor designs, assigns, observes, collaborates in, and interprets the work of students in their classrooms. Additionally, language instructors in general assign meaning to interactions and evaluate the information that they receive and create

in these settings therefore "teachers are the primary agents, not passive consumers, of assessment information. It is their ongoing, formative assessments that primarily influence students' learning." (NCTE – National Council of Teachers of English, 2010) As previously stated, in order to ensure reliability in rubrics, constant teaching training with intensive focus on its practical side (i.e. actual application) is imperative. In other words, it does not matter how "perfect" the rubric for the students needs is if the teacher who will grade with it does not know how to use it soundly; without this vital expertise nothing will work out acceptably during the overall management of this or any other oral exam.

One of the recommendations that Leon and Maldonado (2017, p.63) made after analyzing the results of their research carried out in this Ecuadorian state university was that the language teachers from this Center for Foreign Languages needed training in order to "better comprehend the concept of speaking fluency, the steps, techniques and tasks that are required and aimed at promoting it. This training can come internally, externally or … both." To make a note on this recent suggestion is important for these researchers as it backs up the idea of teacher training for the EFL instructors of this language center in general.

Before conducting it, these researchers also considered other issues related to effective teacher training regarding use of rubrics, more specifically the appropriate use of language, the real English proficiency level of the instructors, the teachers' people skills and their world-knowledge.

The appropriate use of language is a condition to be met by the language tutors before and while using rubrics as it should reflect "subject-specific and curricular knowledge, pedagogical and classroom-management skills, contextual awareness, and an understanding of their students (e.g., Andrew et al., 2005; Andrews, 2003a, 2003b; Çakır & Alici, 2009; Elder, 2001; Hill, Rowan, & Ball, 2005; Troudi, 2005)." (Council of Ministers of Education, Canada (CMEC), 2013, p.16)

The next concern was the real English proficiency level of the instructors, which is "among the most important elements in a teacher's professional repertoire", and it was taken care of by following the recommendations of CMEC (Council of Ministers of Education, Canada, 2013) which demanded revising that the instructors

> possess, along with the disciplinary knowledge required to teach the curriculum, the same language competencies they seek to develop in their students and the procedural and pedagogical competencies required to teach and to support the

development of language skills (Laplante, 2000; Mottet, 2009; Paradis, 2004). (p.16)

The third worry related to the people skills accessible in the teachers as they frequently confront having to instruct dissimilar collections of learners whose traits include diverse stages of abilities and an assortment of "cultural, socioeconomic and linguistic backgrounds." (Council of Ministers of Education, Canada (CMEC), 2013, p.16) Finally, the last apprehension was the instructors' world-knowledge (i.e. "cultural and linguistic diversity [present] in their classrooms") since this factor helps teachers to provide learners "with meaningful learning experiences" that will "enhance their academic achievement and success" (Council of Ministers of Education, Canada (CMEC), 2013, p.16) within their educational context.

The researchers pursued to do the assessment of these students' performance, on this oral exam proposal, in the same manner that is executed for PET. In other words, the candidates were only assessed on their language skills, their development of the conversations and their responses to the tasks. (Downes, 2004, pp.56-57) Additionally, the marks that the teacher awarded to the pupils were "based on performance in the whole test, and [were] not related to performance in particular parts of the test." Finally, although the students do the oral exam in pairs, the performance of each learner is graded individually and all those marks are correctly supported by the suggested rubric, which covers these four "analytical criteria: Grammar and Vocabulary, Discourse Management [i.e. Fluency], Pronunciation and Interactive Communication [i.e. Interaction]." (Downes, 2004, p.57)

At this point, the reader will have a swift summary on what each criterion found in the rubric used to grade this oral exam proposal entailed:

> ***Grammar and Vocabulary*** refers to the accurate and appropriate use of grammatical forms and vocabulary.
>
> ***Discourse Management*** [i.e. Fluency] refers to the candidate's ability to maintain a coherent flow of language …, either within a single utterance or over a string of utterances.
>
> ***Pronunciation*** refers to the candidate's ability to produce comprehensible utterances to [fulfill] the task requirements. This includes stress, rhythm and intonation, as well as individual sounds.
>
> ***Interactive Communication*** [i.e. Interaction] refers to the candidate's ability to use language to achieve meaningful communication. This includes initiating and

> responding without undue hesitation, the ability to use interactive strategies to maintain or repair communication, and sensitivity to the norms of turn-taking. (Downes, 2004, p.57)

Process of standardization for this language center of this oral exam proposal

Up to now, these researchers have reported to the reader on several aspects regarding this oral exam proposal. Those facets include why crafting it for this language center was a necessary action, what characteristics of a good test are included on it, what type of consequences should stem from turning it into the standardized speaking test of this state university, and what speaking components will be considered for it. These social scientists have also indicated how "fluency" was defined as "communicative competence" and "speaking effectiveness" for this oral exam proposal, why and how the skill of speaking should be taught in the classroom, and what types of tests for assessing speaking along with the types of tasks from speaking tests were analyzed for this new verbal assessment scheme. Additionally, these EFL instructors have defined critical thinking as well as the notions behind an ideal critical thinker, presented all fifteen critical thinking skills to the reader and specified how they can be developed in an ordinary classroom, provided specific steps so that any teacher can become an ideal critical thinking instructor, and notified the person reading this dissertation on how the critical thinking skills could be graded with this oral exam proposal.

During this literature review, these researchers have also made emphasis on how the speaking components of "interaction" and "accuracy" were measured in as well as facilitated to the learners for this oral exam proposal, respectively and separately. In addition, these social scientists explained what rubrics are, their main characteristics, and their core components. These EFL instructors have also written about the types of rubrics there are, why rubrics should be implemented to grade the speaking skill, the validity and reliability of rubrics, and finally, how the whole construct of rubrics was utilized during this educational project. Therefore, the time to explain how this oral exam proposal will become the standardized oral progressive English exam for EFL students of Pre-intermediate level in this Center for Foreign Languages of this Ecuadorian state university has come.

Once the proposed spoken test has been analyzed as well as accepted by the administrative and academic authorities of this educational facility, these researchers together with the teaching staff of this language center will have to complete this exact process:

1) Whoever is in charge of designing and printing the material of this oral exam will have to pay close attention to the original format provided by these academics as for a test to be standardized, the "presentation of [this] test should be uniform across forms" (Fulcher & Davidson, 2007, p.120) hence no change should be done to the new format, in any way. This condition is compulsory because when "a test taker opens a paper [i.e. looks at the format, in this case] it should resemble previous forms." (Fulcher & Davidson, 2007, p.120)

2) Since this oral exam proposal includes specific guidelines that have to be followed by both teachers and students while doing and taking it, according to Fulcher and Davidson (2007),

> All [these] instructions should be standardized, and presented in the same font type, style and size. In order for this standardization to be implemented from form to form and year to year, the definition of the appearance of the test needs to be written and stored. Ideally, templates should be created so that new content is simply dropped into a standard layout. Once again, the reason for this standardization is that there is no construct-irrelevant variation that impacts upon test taker performance. (p.120)

Before carrying on to the next step, these EFL instructors advise the reader to have a look at the short script aforementioned (APPENDIX 5) to become aware of the explicit commands for teachers partaking and students involvement that have to be executed during the two turns (i.e. "individual" and "interactive") available on this speaking evaluation.

3) The administration of the tests must also be standardized, and that is done by "doing things consistently, in ways that are prescribed" (Fulcher & Davidson, 2007, p.127). Standardization in test administration is not an easy thing to accomplish, but when there is "consistency in language testing", then "variation[s] in procedure that could threaten score meaning" (Fulcher & Davidson, 2007, p.127) are avoided. Therefore, it is an obvious effort that has to be made.

4) It is compulsory to "control the influence of context through standardization – in essence, [there is an] attempt to *fix* the context – so that each individual experiences

essentially the same test and contexts of administration." (Fulcher & Davidson, 2007, p.198) Once this is achieved, "standardized assessment" (i.e. it does not matter if "the unit of analysis" is a learner's grade or several groups scores, the end results can be associated to the lessons and settings wherein the measurement is used) as well as "comparability of scores" (Fulcher & Davidson, 2007, p.198) are doable commissions for the teachers.

5) The EFL instructors that will be in charge of the test administration are obliged to warrant that "raters or interlocutors in [this] speaking test employ the behaviours that are needed to elicit a ratable speech sample" by simply "making sure that the space for testing is properly organized". (Fulcher & Davidson, 2007, p.258)

6) The training of interlocutors (i.e. "**interlocutor** – a person who engages a candidate in conversation in an oral test, but who does not mark the candidate (that is done by someone else)" – Harmer, 2007, p.276) is a part of the standardization process that shows ample difficulties (i.e. "variation by the interlocutor [is] a potential threat to validity" – Fulcher & Davidson, 2007, p.258) however it is a necessary risk taken in an attempt to have samples of oral exams with "the range of functions that are evidenced in tests where there is an interlocutor present". (Fulcher & Davidson, 2007, p.258)

These six challenging and demanding steps are all that it would take to transform this innovative spoken test in the new standardized oral progressive English exam for the EFL students of Pre-intermediate level enrolled at this Center for Foreign Languages.

What kind of backwash effect took place in this research?

Once all the process that was undergone to produce this oral exam proposal (i.e. choosing the ideal format, creating the activities, drafting the rubrics) and to train these set of learners for it (i.e. developing fluency and emphasizing the use of critical thinking skills) have been meticulously described, these researchers deemed essential to dedicate some lines to the one problem that created a whole sequence of changes regarding the initial preparation these researchers intended to tackle – *the backwash effect*.

Paker (2013, p.1464) united the opinions of several scholars (i.e. Alderson & Wall, 1993; Bachman, 1990; Brown, 2004; Brown & Hudson, 2002; Cheng, 2005; Hughes, 2003; Weir, 1990) to identify backwash effect as the "effect of each test item on teacher's teaching and learners' learning in terms of positive and negative aspects." Alternatively, Harmer (2007, p.167) has defined the "**washback/backwash effect**" as

the condition which arises "when teachers see the form of the test their students are going to have to take and then, as a result, start teaching *for* the test." This definition is similar to the one provided by Prodromou (1995, p.13) who characterized it as "the direct or indirect effect of examinations on teaching methods." In fact, the words of Prodromou may be made certain by Paker (2013, pp.1463-1464) for whom no "matter what we teach in the classroom, … our test items create the needs for our learners to master the knowledge, skill or performance" that will be examined.

The evident backwash effect within reach of this oral exam proposal did not only have negative consequences though, it also presented two positive results: 1) these social scientists became fully aware that to teach speaking as a skill, the context, purpose of communication, the relationships and the roles among the participants are crucial aspects that have to be considered for the meaning of language interactions and the use of language during class. 2) When individuals gain speaking competence, it does not mean that they have acquired full linguistic competence as it entails "the ability to make appropriate use of a variety of language forms (ranging from formal to informal) across the four language modalities: reading, writing, speaking, and listening." (Council of Ministers of Education, Canada (CMEC), 2013, p.16)

To finish this fraction of this literature review, these EFL instructors ask the reader that while reading the rest of this dissertation, keep in mind that they had to forfeit their original plan due to the appearance of the negative backwash effect, right in the middle of the study.

2.2 Primary study

This research was based upon a study originally conducted in Turkey, specifically at the School of Foreign Languages of Pamukkale University Faculty of Education, by Dr. Turan Paker and Ph.D. Devrim Höl. The title of that original study is *"Attitudes and Perceptions of the Students and Instructors towards Testing Speaking Communicatively"* and it was published by the Journal of Education of Pamukkale University, in 2012.

The main goal of that educational project, which could be considered as the starting point for the methodology fraction of this research back in 2016, was to discover the thoughts and insights of the learners and English teachers in relation to the oral assessment in that School of Foreign Languages. The core data collection tools of

that study were two questionnaires, adapted from Güllüoğlu (2004) and constructed on a 5-point Likert scale (Paker & Höl, 2012, p.15), that were delivered to two assortments of partakers.

The first set of feedback forms had forty-four items which were filled out by two hundred and ten students – attending classes for the elementary curriculum of diverse schools and with English proficiency levels varying from pre-intermediate to intermediate – for evaluating their perceptions and attitudes about the verbal communication examination during the whole assessment cycle (i.e. pre-, while and post-tests). From those data sources, whose Cronbach's alpha coefficient was .78 (i.e. its "measure of internal consistency" was "acceptable") (UCLA Institute for Digital Research and Education, 2017 & Deviant, 2017), these are the most important conclusions reached by those students concerning the test and their own performances: "Students pointed out that they could not express themselves adequately during the test, and claimed that they needed to have more oral practice in the classroom." (Paker & Höl, 2012, p.13) Likewise, thanks to the results gotten in that portion of their research, Paker and Höl (2012, p.13) are able to contend that for this particular assembly of learners "the speaking test was regarded as the most difficult test when compared to the testing of other language skills."

The other opinion polls consisted of twenty-six items that asked thirty-two instructors to share their opinions regarding the marking scope, the items employed during that examination, the mechanics (i.e. process) of the evaluation, and the total time allotted for that speaking assessment. According to Paker and Höl (2012, p.13), the information gathered from those questionnaires, whose measure of internal consistency was also acceptable (i.e. Cronbach alpha coefficient of .76), clearly reflects the two following facts: 1) "… the speaking test was the most difficult one to apply and assess" for those teachers. 2) "... the scale and rubrics were adequate enough to assess the students' oral performance", as stated by those language instructors.

As a final point, the conclusions reached by those learners in conjunction with the opinions shared by those EFL teachers are all supported by SPSS (i.e. Statistical Package for the Social Sciences) 16.0, which is in broad terms "general statistical software tailored to the needs of social scientists and the general public." (Blumenthal, 2010, p.1) To conclude this section of the dissertation, it is noteworthy to mention that the main characteristics of SPSS 16.0 are: fine organization and analysis of data; capability of rearranging data, computing new figures and performing an array of

numerical inspections; and permitting simple input/output administration. (Blumenthal, 2010, p.1)

This is the very last point in this literature review consequently the reader must now proceed to the next chapter of this dissertation.

CHAPTER 3

3. RESEARCH PARADIGM

In this chapter, the reader will be informed on some essential details that permitted this pair of social scientists not only to discover the impact that these undergraduates bore during the experience of being correctly assessed on the productive skill of speaking, but to also portray the perceptions that they shared after that suitable evaluation.

The following points will be addressed in the *methodology* segment of this educational assignment: 1) the reader will be notified on the ontological as well as the epistemological perspective shared by these researchers all throughout this final task. 2) Some information about the sampling technique employed, the two principles concerning effective assessment that were followed (i.e. validity and reliability), and the reasons why there is no hypothesis to be tested on this research project will be delineated on the reader's behalf. 3) The reader will be enlightened on the statistical significance test brought into play to transform this dissertation into a potential and trustworthy reference regarding this issue.

For the *nature of data* section, these researchers will name the original source they stood by and the procedures they adhered to all along the main events of data gathering for this research. In the fraction named *features of data analysis*, this pair of EFL instructors specifies the analytics trial that they selected as appropriate for the needs of this study along with the different mathematical elements which are part of that statistical examination.

The other three imperative pieces of information presented in this chapter are: 1) description of study location, 2) portrayal of participants, and 3) ethical considerations. These final components are central and crucial for any reader to be able to grasp the whole picture of the research paradigm wherein this study was built upon.

3.1 Methodology

Ontologically speaking, this pair of researchers is adhering to the constructivism paradigm as this particular project will examine and analyze an

explicit set of social phenomena, along with their meanings, that is "continually being accomplished by social actors" as stated by Bryman (2001, pp.16-18) in Grix (2002, p.177). These writers also support their decision of applying a constructive perspective on this research due to the conceptions provided by Dudovskyi (2016) who directly quotes Elkind (2005) to state that constructivism is "the recognition that reality is a product of human intelligence interacting with experience in the real world" hence "reality is perceived to be subjective" in this ontological paradigm for the former author.

On the epistemological flank, these researchers are following the interpretivism perspective because the attributes pertaining to this particular project ask for a strategy "that respects the differences between people and the objects of the natural sciences and therefore requires the social scientist to grasp the subjective meaning of social action" in accordance with Grix (2002, p.178) based on Bryman (2001, pp.12-13). Another reason for which this interpretive paradigm will be employed is that it will allow these language instructors to "understand and interpret the world in terms of its actors." (Cohen, Manion, & Morrison, 2007, p.26)

For this research, all the participants will be selected using a simple random sampling as "each member of the population under study has an equal chance of being selected and the probability of a member of the population being selected is unaffected by the selection of other members of the population," (Cohen et al., 2007, p.110) which means that each selection is entirely independent of the next. Following the constraints proposed by Cohen et al. (2007, p.111), this sampling method is the most suitable for this study as the resulting sample (Original Sample Size: 70 learners) will be comprised by subjects with similar characteristics (i.e. EFL learners, Spanish speakers, Ecuadorians, state university undergraduates, young adults) to the population as a whole.

Validity is an important issue that has to be considered while drafting any kind of research. For this individual case, and in order to demonstrate and verify validity on this project, these researchers will call upon two out of the five kinds of validity argued by Maxwell (1992), as quoted by Cohen, Manion and Morrison (2007, p.135): *"descriptive validity"* and *"interpretative validity"*. Firstly, this research will have descriptive validity as these EFL instructors will narrate the facts that occur during the project in an accurate manner, without making, selecting or distorting anything up. Secondly, this project will also possess interpretive validity

since the design of this research will allow these EFL instructors "to catch the meaning, interpretations, terms, intentions that situations and events, i.e. data, have for the participants/subjects themselves, in their terms."

In spite of the presence of those validity items, "**internal validity**" (Mackey & Gass, 2005, p.109) could be a worrisome issue attributable to participant attitude due to the fact that all seventy participants (i.e. Original Sample Size) of this research will be chosen - even if it is randomly - from four different groups. It could occur that those selected undergraduates become prey of the Hawthorne effect (Mackey & Gass, 2005, p.114) as they may possibly feel that they are "different" compared to their unselected peers simply because they are part of the experiment, which might end in incorrect or careless participant behaviour.

For reliability purposes, this study will follow the concept behind *"inter-rater reliability"* as depicted by Denzin and Lincoln (1994) in Cohen et al. (2007, p.148). The former scholars argued that reliability can be addressed as replicability in qualitative research, if it merely endorses certain means, such as stability of observations, parallel forms, and inter-rater reliability. As this project asks for the participation of other EFL instructors who work at the same language center, who will be trained on the proper use of rubrics, and who will actively part-take in the collection of the data, they should be able to corroborate, support and duplicate the results obtained by the authors of this research in their own classrooms – inter-rater reliability.

There is no hypothesis to be tested on this research project as its focus will be qualitative. The authors of this proposal have decided to follow this approach as it is the best one for this specific project since its particular characteristics assert it that way.

To fully understand the preceding pronouncement, it is necessary to resource to Creswell (2015, p.16) for whom the main features of qualitative research include

1. Exploring a problem and developing a detailed understanding of a central phenomenon.
2. Having literature review play a minor role but justify the problem.
3. Stating the purpose and research questions in an open-ended way to capture the participants' experiences, collecting data based on words (e.g., from interviews) or from images (e.g., photographs) from a small number of individuals so participants' views are obtained.
4. Analyzing the data for description and themes using text analysis and interpreting the larger meaning of the findings.

5. Writing the report using flexible, emerging structures and evaluate criteria and including the researchers' subjective **reflexivity** and bias.

Based on those notions, this pair of researchers acknowledges and has been able to prove that a hypothesis is not a compulsory fraction of this approach.

Nonetheless, for making the gathered data, obtained results, and reached conclusions more accurate, much more important, and a great deal more significant to the Ecuadorian EFL teaching community, these social scientists decided to prove the null hypothesis that corresponds to this study via one of "the most commonly used statistical **significance tests**" (Efstathiou, 2012): "**Student's t-test**".

In this case, the null hypothesis – whose purpose is to "test in the general population that there is no change, no relationship, or no difference", according to Creswell (2015, p.126) – states that there is no difference in the development of the speaking skill of undergraduates with B1 (Pre-Intermediate) English proficiency level at an Ecuadorian state university, once this intervention has been concluded. Keep in mind that this null hypothesis can be extrapolated from one of the specific objectives of this research (i.e. #3), available on the "statement of the problem" section in Chapter 1.

It is essential to report that this extensive analytics trial (i.e. student's t-test) was performed with the intention of turning this whole study into an article worthy of being cited in the future hence this pair of researchers had to put these figures under statistical scrutiny so that its valuable observations might be impartially provable and arithmetically verifiable.

3.2 Nature of data

A structured questionnaire adapted from a study done by Paker and Höl (2012) – **"Attitudes and Perceptions of the Students and Instructors towards Testing Speaking Communicatively"** – was designed to collect data from students (APPENDIX 1). The questionnaire to be administered to students has twelve items, which will find out their perceptions and attitudes towards this standardized oral exam as well as their experience before, during and after the pre- and post-exam.

The data to be collected during this study will be gathered in two different stages: Stage One involves the use of two tests, a pre-test and a post-test. Stage Two entails filling out a survey with a structured questionnaire that will be portrayed as Students' Interviews.

The pre-test will be taken by those students previously selected (Original Sample Size: 70) and it will consist of an oral exam based on the present format. This verbal assessment will be directed by the chosen teachers during the mid-term oral exam, which takes place during the first term. This pre-test (APPENDIX 3) will be graded using an analytic rubric (APPENDIX 4) and those scores will be used as the first data.

The post-test (APPENDIX 5) will be taken by the same students previously selected (Original Sample Size: 70), but its format will follow the guidelines available on the proposal of these researchers for a standardized oral exam. This spoken evaluation will be directed by the same chosen teachers however they will switch classes for this final oral exam. That post-test will be graded using the proposed rubric, an adaptation of the previous ones which had been applied in the pre-test, and those scores will be considered as second data.

Between both exams, pre and post tests, the five weeks' intervention will occur. Teachers will be trained and "calibrated" in the use of the new rubric (APPENDIX 6) to evaluate the oral proficiency of these students applying a standardized procedure. Also, as part of students' instruction and preparation, teachers will be asked to apply speaking strategies and activities with their English classes following some guidelines to improve those students' oral production.

3.3 Features of data analysis

Analytics will be used to interpret and compare results, using the first and second data collected in mid-term and final oral exams as inputs, to observe if there was any difference in those outcomes.

These researchers will use two groups, with thirty-five undergraduates (Original Sample Size: 70) each. During the time span of the research, these students will receive the same treatment. Besides, at the end of the post-exam (i.e. final oral exam), they will be asked to answer a structured questionnaire to understand the impact regarding the treatment used in class, and to comprehend their perceptions concerning the standardized oral exam.

The statistical significance test that this pair of social scientists chose to analyze the inputs previously mentioned is the student's t-test. This technique allows you to compare the results and verify the hypothesis of the study. In this case, as two

measurements are obtained (i.e. x-prior and x-post) in the same group, this test of statistical significance is applied to show the difference of the arithmetic means attained in the research. (APPENDIX 11)

3.4 Description of study location

This research came about in the premises of an Ecuadorian state university, more specifically at the facilities of its Center for Foreign Languages. Therefore, these two places will have to be described with detail in the next lines.

The Ecuadorian state university where this study took place is one of the largest of its kind in the country. This educational establishment, with sixty years of teaching experience and didactic development in the Ecuadorian territory, proposes several degrees in diverse professional fields (i.e. tourism, robotics, mechanical engineering, graphic design, engineering - the most chosen one - and economics) to its approximately nine thousand enrolled students.

One of its nationwide finest achievements occurred in 2009 when it was certified as the best Ecuadorian university by CONEA (Consejo Nacional de Evaluación y Acreditación de la Educación Superior del Ecuador) – the national accreditation board at that time. The Ecuadorian regime considered CONEA's authorized endorsement as the central foundation for its decision to call upon the extensive expertise that this state university has in order to lead it and collaborate on some undertakings related to the development and enhancement of the English language proficiency area in the country. It is compulsory to mention that this institution of higher education made use of its Center for Foreign Languages in order to carry out the instructive errand explained above.

Between 2011 and 2013, the Ecuadorian national government through CEACES (Consejo de Evaluación, Acreditación y Aseguramiento de la Calidad de la Educación Superior) - the present national accreditation board - evaluated and appraised all the universities and colleges of the country. The main objective of this endeavor was to measure the quality of education that Ecuadorian undergraduates were receiving back then, with the hidden intention of shutting down those educational institutions which did not meet up certain specific criteria. It is noteworthy to point out that this "cleansing" plan was done for all the syllabi which were taught at those now non-existing scholastic establishments, not only for the English area.

Under direct commission of the Ecuadorian Department of Education, the Center for Foreign Languages of this state university had to evaluate and determine the English skillfulness level of the students that attended those universities and colleges aforementioned. With these goals in mind, the staff of EFL instructors that was working for this language center created, invigilated and assessed a batch of exams whose grades clearly reflected that the learners had or had not reached the English language prerequisite before graduating from their degree. If that educational law criterion was not accomplished by every single one of those undergraduates, then those students proceeded to do another exam, which was also graded by the staff of this language center turning that examination into a mandatory graduation requirement, basically all over the country. It must be stated that this project has been the largest one that this Center for Foreign Languages has handled up until now, but it has not been the only one designed, developed and elaborated for the Ecuadorian government.

For a period of two years since 2014, the Center for Foreign Languages of this state university worked with the legislative government of Ecuador to supply its collaborators (i.e. employees of the public sector) with preparation courses for TOEFL (Test of English as a Foreign Language) and GRE (Graduate Record Examination) so that they could attain the required level of improvement in this second language necessary to apply, complete and excel at master degrees as well as doctorate programs in foreign countries where Spanish is not the preferred language of instruction.

This state university does not only focus on the external clients (i.e. Ecuador's government and society), it also cares deeply about its internal customer (i.e. its own undergraduates and teaching/learning community). English is a requirement in the curriculum of the students registered in this state university, and if these learners hope to get graduated, they have to exhibit and prove a "**B1**" English level. In other words, they must be an "**English Independent User**" as described by CEFR (Common European Framework of Reference for Languages). Thus, these learners should be able to do well and pass Cambridge Preliminary (PET) and BEC Preliminary. They also ought to be capable of getting scores from 40 to 59 in BULATS, from 57 to 86 in English TOEFL, and reach Level 4 in IELTS. (English Proficiency Test Online & Express Publishing, 2018)

The Center for Foreign Languages of this state university has been undergoing significant and critical changes since 2005. Such transformation, in both the academic aspect and the physical realm of this language center, was not only inescapable, it was

imperative as the approximately four thousand students it holds in the nearly one hundred courses it has at any given semester – as a rule – demanded it.

In the first sense, the continuous preparation that the specialists of this foreign language department have been receiving for the last thirteen years – either "in-house" or otherwise – is clear proof that its academic authorities pursue excellence among its teachers. These permanent training sessions asked the staff to form an opinion on their teaching practices and habits just to come to the realization that there are always better, more engaging and more interesting ways to *grow* professionally, to be better teachers for their students (thirty-five to forty-five learners in any 'common class' of this state university), and to gain access to their personal aims more rapidly. On the second factor, this language center received its own office block and classrooms in 2011. These material resources have allowed its teaching staff to develop their class sessions in spaces which hold the appropriate technological and ergonomic characteristics that an up-to-date, high-quality, suitable EFL classroom must possess. Before having its own premises, the language instructors of this language center used to teach in borrowed classrooms of this state university's schools whose conditions were neither adequate nor useful for EFL teaching and learning.

To come to an end with the description of the location wherein this research was developed, some final comments regarding this state university: CEACES sized it up in 2013 and it was labeled with the highest rank – Category "A"– once again. In fact, this public institution of higher education is still found and favored at "**CATEGORÍA A**" level nowadays, which means it upholds CEACES's utmost official approval. (educarplus.com, 2017)

This pair of researchers chose the Center for Foreign Languages of this state university as the place to be employed for this study due to the scholastic, tangible, and fitting elements that were within their grip. Those advantageous components consisted of having steady entrance to its facilities as well as prompt contact with those learners who were willing and able to participate in this research. Furthermore, these social scientists observed and reckoned that this educative investigation would definitely exploit the underlying and solid mind-set to train for accomplishing academic feats available inside those undergraduates.

3.5 Portrayal of participants

Now that this study location has been described, it is time to make a comprehensive portrayal of the learners involved in this research. The distinctive features present in this set of students in addition to their general attributes will be depicted next.

The common traits that these learners shared included age, gender, motivation for attending classes at this state university, future expectations within their professional fields, their prospects in the educational side, and their hopes for a better life. The typical undergraduate of this state university is a young adult (17 or 18 to 21 or 22) who is studying here because he (there are more male than female students in this public institution of higher education) is interested in completing and getting hold of a university degree (formal education instead of technical realization), which will let him work at a multinational company (a secure job over entrepreneurial ventures). This student also plans to do a masters degree abroad (usually from English speaking countries) right after graduating from this state university, as he is aware that he will only be compensated with a superior standard of living (a nicer future in general terms) if he has attained a higher academic training during his youthful years.

On other respects, these undergraduates shared the following general attributes, some of which were complex to deal with as well as adverse for the natural progress of this research: the same nationality (all the learners involved in this project were Ecuadorian men and women as there was no foreign partaker in this research), similar social status (they live in households with incomes that are significantly lower than the middle class homes), identical educational background (these pupils graduated from high schools where achieving high proficiency in the English language is not a priority), equivalent phases of completion for their corresponding curriculums (they were all in the mid-part of their degrees), equal time of exposure to the teaching methodology of this Center for Foreign Languages (it was not the first course in this language center for any of the participants), and different professional interests (the constituents of this assembly were from assorted degrees).

Linguistically speaking, the most distinctive features that were allocated in this collection of students comprised these learners' monolingualism (Spanish as their L1 – First Language) and their real English proficiency level (an array of mixed stages, which illustrated that many of them were "Basic Users" – either "Beginners" / A1 or

"Elementary" / A2 – and some others were barely grasping the "Independent Users" echelon – "Intermediate" / B1). (English Proficiency Test Online & Express Publishing, 2018) This last point is certainly a serious issue given that it is not the phase these undergraduates should be tackling and dealing with at this time as they are supposed to be well on their way to master B1, not just on the verge of accomplishing it. What is even worse is that more than a few of the pupils from this intervened cluster were learners who were only attending these English classes since it was a requirement of their degree programs, not because they perceived this educational experience as an opportunity to enhance their understanding of this second language or simply as a learning prospect.

3.6 Ethical considerations

Before the research is initiated, the participants will have to read and sign an informed consent form that will follow the codes published by the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research in 1979. Those policies are also known as The Belmont Principles and they regard truly important aspects, such as beneficence/non-maleficence, respect for persons/autonomy, and justice. After the participants have gained "fully informed consent where appropriate", this pair of researchers will have to confirm that they know they "have the right to withdraw at any time". Additionally, these researchers will have to "inform participants [on] who will have access to the data/report, how public it will be, when it will become public and how it will be disseminated." Finally, this pair of social scientists will have to "negotiate levels of release" and most importantly, these researchers will have to "ensure anonymity/confidentiality/non-traceability." (Cohen et al., 2007, p.77)

CHAPTER 4

4. GENERAL INFORMATION ABOUT THE STUDY AND DATA ANALYSIS

In this chapter, the reader will find the initial instructional design of this research, which later on had to be adjusted due to circumstances beyond the reach of these researchers. The reader will also become aware of the intricacies belonging to this study as well as its limitations in this section of the dissertation. The amendments made to the original coaching plan of this investigation are carefully explained in this chapter too. To finish this part of the thesis, this exploration produced qualitative information as well as quantitative data that will first be presented to the reader, and after that, these records and figures will both be analyzed in order to be properly interpreted by these academics.

4.1 Initial instructional design of the study

The original coaching plan of this investigation was proposed, devised and set by the Critical Thinking Expert whose instructional design involved practicing Critical Thinking activities, which required hands-on involvement of these learners, in the class sittings that would come about during this intervention. It is compulsory to emphasize that these tasks were not only going to develop the Critical Thinking skills of these students, they were also going to cover the compulsory syllabus of the language center where this study was applied.

The activities aforementioned concentrated their attention in meaningful learning, which according to Nation and Newton (2009, p.19) means that these tasks had to "focus on meaningful and relevant content". The Critical Thinking Expert made this decision after considering the "five principles for teaching beginners" that Nation and Newton (2009, p.19) put forward for consideration in their book called Teaching ESL/EFL listening and speaking. Those five principles, which are "particularly relevant to the teaching of beginners" (Nation & Newton, 2009, p.19), will be listed next so the reader can go through each one of them:

1 **M**eaning	Focus on meaningful and relevant language
2 **I**nterest	Maintain interest through a variety of activities
3 **N**ew language	Avoid overloading learners with too much new language

4 **U**nderstanding Provide plenty of comprehensive input

5 **S**tress-free Create a friendly, safe, cooperative classroom environment

(Nation & Newton, 2009, p.19)

Apart from giving consideration to the "MINUS" principles of Nation and Newton (2009, p.19), the Critical Thinking Expert mulled over the "activity types" that according to Williams and Puchta help to "develop thinking skills and language". (Puchta, 2012, pp.16-17) These scholars built up "13 categories of activity", which "roughly follow a cline from basic to higher-order thinking skills" (Puchta, 2012, p.17), and they are: "making comparisons, categorising, sequencing, focusing attention, memorising, exploring space, exploring time, exploring numbers, making associations, analyzing cause and effect, making decisions, solving problems, [and] creative thinking". (Puchta, 2012, p.17) The Critical Thinking Expert took advantage of a number of these activities for the instructional design of this study and the reader must be conscious of that fact.

Finally, the Critical Thinking Expert worked under the premise that meaningful activities help all learners improve their performance, productivity and output when attempting to develop their productive skills. As a result, since speaking is regarded as a productive skill, this set of activities would provide these students with plenty of useful knowledge. Moreover, these tasks would become valuable tools that might have allowed them to improve their upcoming oral examinations scores.

4.2 Intricacies of the study

Regarding the pivotal decision of settling on which approach to utilize for introducing specific aspects of the English language (syntax rules + lexical exponents) to these two groups (i.e. Original Sample Size: 70 participants), these researchers opted to employ the focus they were most accustomed to practicing during the classes they had previously taught wherein the main goal was developing the speaking skill. These EFL instructors proceeded this way as those dissimilar designs had been quite successful in prior courses, within their corresponding classrooms. Therefore, the Critical Thinking Expert selected to yield an "***inductive approach***", as depicted by Harmer (2007, p.82), during his class sessions – i.e. the students were exposed to examples of the Target Language and figured out the rules by themselves to be exerted during an ordinary spoken interaction. Contrariwise to that brand, the Experienced Oral

Examiner for International Exams exercised a "***deductive approach***", in accordance with the portrayal of Harmer (2007, p.81), all throughout the training hours of his group – i.e. the students were given explanations about grammar rules and new vocabulary items to be applied through the lesson before they started producing utterances and using them to express their own ideas and opinions orally.

In the first oral exam, each language instructor was in charge of their own students, which means they graded their own groups. For the second oral exam, expecting to comply with the purposes for validity and reliability of this research (i.e. face validity and marking reliability), the instructors interchanged classes in order not to be biased – positively or negatively – towards the real performance, actual linguistic competence, and existent proficiency level of their learners.

The main intention behind the decision of interchanging groups was to prove that the new format could be applied, fairly and effectively, independently of the language instructor in charge of completing this oral examination with these learners. Furthermore, the proposal itself was developed considering the fact that both the format of the activities and the rubric to be used for grading them were well-defined, comprehensible and concise hence uncomplicated to handle as well as equitable and justifiable while employed. It is noteworthy to indicate that even though they were not the ones grading their performances, both teachers were present in the classrooms where these spoken examinations took place, ushering and accompanying their pupils all throughout the oral exam process of these collections of learners so as to lower their anxiety levels – as much as possible – and for easing them up into this pioneering occurrence.

The entry exam model (APPENDIX 3) as well as the exit test format (APPENDIX 5), along with their internal processes, were going to permit these researchers compilation of accurate and measurable data during the oral examinations of these undergraduates. This reliable information would consequently allow these EFL instructors to linger on and elaborate upon the intervention plan right before this study was initiated, and also well after this research was concluded.

To finish writing about the complexities of this study, the reader has to know that its quantitative information was obtained by means of the same entry and exit exams mentioned above. For the first examination, this data was calculated by applying the existing rubric (APPENDIX 4) used for the current verbal assessment. Alternatively, for the second one (i.e. this oral exam proposal), these academics assessed and

evaluated the spoken performances of these students through a structured (i.e. analytic) rubric – with balanced measures for "Grammar and Vocabulary", "Fluency", "Pronunciation", and "Interactive Communication" – of their own creation.

4.3 Study limitations

This study had five easily recognizable stages (i.e. general planning; study preparation and completion; results analysis; attaining deductions and reaching conclusions; future plan of action) which required the direct and constant intervention of these researchers. As a matter of fact, the first two instructive junctures of those five phases had been thoroughly organized, advanced in a timely fashion, and concluded without delay, according to the chronogram scheduled in 2016. These instructional chapters were possible for the main reason that for everything concerning class training, and related to students' performance measurements, both of these social scientists were working together and at the same time in the language center of this state university. Unfortunately, this couple of facts changed during the analysis of results stage due to budgetary reasons of this public institution, which meant the separation of the Critical Thinking Expert from this language center as he had only been working there for six months. Even though this unfortunate setback has nothing to do with the actual materialization of the initial plan, it serves the purpose of explaining why these figures were not analyzed on time or within the fixed schedule, as it had been originally intended. For that reason, the first serious limitation that this study will have to face has to do with the fact that its data may be regarded as outdated instead of enduring.

Another possible deterrent for this research will definitely be the authentic English proficiency level that several students of this whole group have as it might create doubts on the learners minds about whether or not they are really capable of undertaking, and actually doing, the sort of training (i.e. tasks to improve their speaking competence and projects to develop their critical thinking skills) needed to succeed, or at least perform well, in this educational project.

Other substantial constraint for this proposal of a standardized oral exam is being required to follow a predetermined study program, constantly and strictly. Consequently, not too many opportunities for academic freedom exist, nor a lot of space for experimenting with new methodologies either, unless they have been set by the Language Center direction, the Language Center academic coordination or as a result of

the bonding of both wills, as it usually is. If the process is not done as preset by these authorities, there is the chance or risk that those students who are not up to standards, whose linguistic basis or proficiency level is not appropriate for the course, will fail the subject and will not advance on the syllabus of their degree.

Yet another huge limitation that this educational project will play against is the occurrence of the incident known as *"washback"*. (Tsagari, 2007, pp.3-5) This phenomenon can be described as the effect that an evaluation, mostly a future assessment, has over the manner a class is taught and revised as it affects the behavior of both the teacher and the students. Unless there is a test whose significance is greater for the undergraduates, this proposal of a standardized oral exam might direct the teaching practices applied by the language instructor and the focus of the class sessions towards what the test, evaluation, assessment requires from the learners so that they get a good grade on it or just pass it.

4.4 Adjustments on the initial instructional design of the study

In the original study that was used as one of the foundations for this educational project, Paker and Höl (2012, p.18) acknowledged the concept of " 'backwash effect' " (Alderson, Clapham & Wall, 1995; Hughes, 2003) as the reason why the learners of their research will "get ready much better in their following speaking tests" and the major cause for which they assured that "testing speaking will be an important component in their (i.e. those students) agenda from now on". This effect, which can also be referred to as the washback effect, is "the influence that a test has on the way students are taught (e.g. the teaching mirrors the test because teachers want their students to pass)" (ELT World Wiki, 2017) hence this induced stimulus will certainly produce changes in both the teaching practice and the learning process. Furthermore, the actual results of any performance assessment or proficiency level numerical estimation which came after that affected and manipulated teaching can leave the learners greatly motivated (i.e. "Positive washback" effect as "there is harmony between the teaching and the students' examination or a class test performance" – ELT World Wiki, 2017) or worryingly demotivated (i.e. "Negative washback effect" since "there is no sync between what is taught and what is performed" – ELT World Wiki, 2017).

For this research, the activities, strategies, tasks and techniques implemented for teaching (i.e. developing and practicing) the speaking skill had to be shaped differently

compared to the ones normally used in the classes of this Center for Foreign Languages as the format utilized in the oral exam proposal compelled it, thus the existence of the backwash effect in this part of the study is substantiated.

The new verbal assessment design consists of this pair of comprehensible and responsive tasks: 1) an "individual turn" that asks the learners to show fluency (i.e. Speaking Competence) while doing the first spoken duty and 2) an "interactive turn" which centers the students attention on providing evidence for problem-solving expertise (i.e. Critical Thinking Skills) during the second oral undertaking. Unfortunately, those communicative projects and spoken endeavors that had already been planned by the Critical Thinking Expert (i.e. RESEARCHER "A") could not be done nor put into operation with the set of learners under the supervision of the Experienced Oral Examiner for International Exams (i.e. RESEARCHER "B") as these last participants became prey of fear of failing future courses due to their real English proficiency level. Therefore, these undergraduates requested this teacher to focus his classes much more on improving the grammatical and lexical knowledge required to pass the level, and a lot less on developing and practicing their speaking skill in general. As this teacher had not considered this sort of predicament, he talked to the other researcher and they both decided that these undergraduates should at least work on expanding and perfecting their speaking competence as the projects which resorted to acquiring and enhancing critical thinking skills were rather heavy on applying prior language rules accurately, recognizing syntax background in different sentence structures, and knowing previous vocabulary well.

In order to assist these students in the development and practice of their speaking competence (i.e. fluency), the Experienced Oral Examiner for International Exams had to consider "the main characteristics of activities designed to develop fluency" – represented in the words of Nation and Newton (2009, p.152) – before drafting and modeling the tasks that they would have to do during their class sessions therefore those activities and tasks should not have only been message-focused and easy, they had to produce high-leveled performances. (Nation & Newton, 2009, pp.151-152) In fact, when those explicit points aforementioned are not present in any speaking activity, it can be stated that the learning of the language items exposed and analyzed during that communication task has been for little purpose if the students do not increase their speed of access to them, if the learners production does not augment much, and if the

number of hesitations pupils make does not diminish; in other words, if fluency has not been achieved. (Nation & Newton, 2009, pp.152-153)

Nation and Newton (2009, p.153) also pointed out that fluency activities rely on "several design requirements and features" to accomplish their mission. The design requirements that those authors alluded to are "easy tasks" and "message focus" (Nation & Newton, 2009, p.154) whereas "time pressure", "planning and preparation", and "repetition" (Nation & Newton, 2009, pp.154-155) are the conceptions conveyed as design features.

When the "learners take part in activities where all the language items are within their previous experience" (Nation & Newton, 2009, pp.152-153) then the "easy tasks" design requirement has been accomplished; however, for the "message focus" design constraint to be met, the exclusive condition is that the "activity is meaning-focused". (Nation & Newton, 2009, p.152) More to the point, the only attributes that fulfill the specifications a fluency developing activity has to draw upon and seize concerning design features (i.e. "time pressure", "planning and preparation", and "repetition") are "support and encouragement for the learner to perform at a higher than normal level". (Nation & Newton, 2009, p.153) These are all the reasons why RESEARCHER "B" took these ideas into account whilst planning the tasks and activities for expanding the fluency stored within this set of undergraduates.

The result of such considerations was that the fluency activities concocted for this cluster of students lasted no more than ten minutes each – including preparation time, had to be done during class – at least twice per class day, and basically followed the same steps. Those straightforward and repetitive guidelines were: 1) starting the interaction with a general question, and 2) making the students continue the conversation by asking each other more questions (i.e. minimum number, two; maximum, five) that were directly related to the answer of the first query.

When the learners did the second step well, they did not just rehearse generating appropriate yes / no questions, they attempted the production of long wh-questions – an action that tends to be much more demanding grammatically and far more complex in the lexical department than asking short questions which can be answered monosyllabically – as the established procedure of the task encouraged them to and required it from them in order to keep the interaction going.

As it can be seen, these activities were easy tasks and meaning-focused, but did they produce high-leveled productions? They did, indeed. After these undergraduates

had done these fluency tasks, they had loosened their tongues and they had gained some confidence on their speaking skill. Those progressions would eventually lead them to develop a certain degree of oral fluency, which was based on the topics that had to be done during their course. Likewise, the content of these activities was always available to the learners as it could be found in each of the units that they had to do for their end-of-term exam. However, this improvement should not be considered significant for the level they are supposed to be, which is much higher.

It is pertinent to state that RESEARCHER "B" also reflected on using the "4/3/2 technique devised by Maurice (1983)" (Nation & Newton, 2009, p.153) however actual class time was a large limitation because when this technique is taken up, it entails whole class movement with loads of interaction bits and these learners were not that cooperative in this regard.

4.5 Presentation of data

The data of this study will be organized in qualitative information (i.e. obtained from the structured questionnaires) and quantitative data (i.e. collected from the numerical results gotten from applying the oral examinations to students). Hence, these records and these figures will be presented fittingly as to comply with the research questions and sub-research questions proposed, and to provide conclusions in accordance with the objectives mentioned in this study.

4.6 Analysis of data from questionnaires

The size of this original sample was composed by seventy (70) undergraduates enrolled and attending classes in a Pre-Intermediate course at the Foreign Language Center of this Ecuadorian state university. As it is foreseeable, the sample size suffered changes due to "experimental attrition" (Lund Research Ltd., 2012), which in this case decreased the initial number to sixty-seven (67) as the "experimental mortality" (Lund Research Ltd., 2012) was three (3) participants in total. The reader must be aware that thirty-five (35) of those sixty-seven (67) students belonged to RESEARCHER "A" collection and the remaining thirty-two (32) to the RESEARCHER "B" assembly.

In order to gather the opinions of these learners regarding the format employed to measure their productive skill of speaking, this group of participants completed two

(2) structured questionnaires: an entry questionnaire and an exit questionnaire. It is important to let the reader know that these groupings of queries (i.e. entry and exit questionnaires) were given to the learners in Spanish – their L1 – so that they could truly appreciate the matter of the questions they were about to answer. Likewise, these inquiries were also identical in form and scope, as they both contained the exact same questions. The entry questionnaire was done right after the whole group had finished doing their first oral exam (i.e. mid-term), which was planned to apply the actual format that is used for this specific "summative assessment" (University of Exeter, 2008) at the Center of Languages that belongs to this educational establishment. Likewise, the exit questionnaire was done right after this set of students had finished doing their second oral exam (i.e. end of course), with the only difference that this "**achievement test**" (Roa, 2014) was based on the proposal of these researchers.

Before starting the analysis of the information provided by these questionnaires, it is key and relevant to remind the reader that both these surveys had the same twelve (12) questions whose main objective was to allow these university students to voice their viewpoints framed upon the options available there. Accordingly, these researchers request the reader to bear in mind that this array of queries was pieced together with the purpose of letting students share their honest and personal opinions (i.e. the actual basis for the qualitative analysis of this dissertation) regarding this explicit educational endeavor with these EFL instructors.

From the methodological point of view, these social scientists produced this structured questionnaire with the focal aim of collecting measurable and analyzable data from these students to provide responses to the research questions as well as the sub-research questions of this study. Consequently, these twelve (12) questions were united under the following combinations:

1) Advantages: Q2, Q3, Q5, Q8

– Here are the corresponding texts of those questions. Q2: "Estaba informado acerca del procedimiento del examen." (I was informed about the procedure of the test.) Q3: "Estaba muy cómodo antes del examen." (I was very comfortable before the test.) Q5: "Fui preparado en clases (a través de estrategias y actividades) para el examen oral." (I was prepared in class (through strategies and activities) for the speaking test.) Q8: "Fui alentado y motivado por el profesor durante el examen." (I was encouraged and motivated by the teacher during the test.)

2) Disadvantages: Q1, Q4, Q6, Q7, Q9

– Next, the reader will find the contents of those queries. Q1: “Estaba más ansioso y nervioso comparado a antes del examen oral.” (I was more anxious and nervous compared to before doing the speaking test.) Q4: “No tenía idea alguna acerca de la escala de evaluación (rúbrica).” (I did not have any idea about the assessment scale (rubric).) Q6: “Tuve dificultad para expresarme claramente y en una forma adecuada durante el examen.” (I had difficulty in expressing myself clearly and in an adequate way during the test.) Q7: “Estaba ansioso y nervioso durante el examen oral.” (I was anxious and nervous during the speaking test.) Q9: “Las tareas/actividades en el examen fueron de tipo que nunca antes había encontrado.” (The tasks/activities in the test were types I had never encountered before.)

3) Recommendations: Q10, Q11

– To keep informing the reader about the substance of these inquiries, here are the subject matters of questions 10 and 11. Q10: “Creo que las actividades orales en clases fueron insuficientes.” (I believe that the speaking activities in classes were insufficient.) Q11: “Después del examen, me di cuenta que necesitaba más practica oral.” (After the test, I realized that I needed more speaking practice.)

4) Approval: Q12

– The essence of the last question is Q12: “Me sentí más cómodo haciendo este nuevo examen oral en vez de los anteriores.” (I felt more comfortable doing this new oral exam rather than previous ones.)

Now that the sources utilized to acquire the data which will be depicted in the subsequent lines have been recalled as well as the valuable facts and details that this information will facilitate to portray, it is pertinent and compulsory to start the descriptive analysis of said information.

In order to present an effective summary of the judgments collected alongside the results gotten during this research, these academics have decided to represent the gathered data in four (4) singular sets as concatenating these figures seems to be the most sensible decision apropos inputs analysis. The aforementioned categories are 1) “Advantages”, 2) “Disadvantages”, 3) “Recommendations”, and 4) “Level of Approval” of using this oral exam proposal. The reader must consider that the “entry results” shown here will be from the “current oral exam format” used in this language center, and the “exit results” explained in the next lines will be related to the “new verbal assessment scheme” proposed by these social scientists.

Let us start with the presentation of the data gathered for this study by comparing the data of the entry questionnaire (i.e. completed in December 20th of 2016 and answered by sixty-seven (67) participants) to the information from the exit questionnaire (i.e. done in February 14th of 2017 and filled out by sixty-six (66) study partakers).

Table 4.1 Entry Questionnaire in English (2016)

ENTRY QUESTIONNAIRE: 67 PARTICIPANTS / December 20, 2016						
Number	Questions	Agree	Partially Agree	Disagree	Completely Disagree	Total
1	I was more anxious and nervous compared to before doing the speaking test.	20	31	12	4	67
2	I was informed about the procedure of the test.	38	21	5	3	67
3	I was very comfortable before the test.	13	27	17	7	64
4	I did not have any idea about the assessment scale (rubric).	7	10	24	26	67
5	I was prepared in class (through strategies and activities) for the speaking test.	34	20	12	1	67
6	I had difficulty in expressing myself clearly and in an adequate way during the test.	21	30	12	4	67
7	I was anxious and nervous during the speaking test.	27	24	13	3	67
8	I was encouraged and motivated by the teacher during the test.	25	27	10	4	66
9	The tasks/activities in the test were types I had never encountered before.	1	20	27	19	67
10	I believe thet the speaking activities in classes were insufficient.	5	15	28	19	67
11	After the test, I realized that I needed more speaking practice.	44	18	3	2	67
12	I felt more comfortable doing this new oral exam rather than previous one.	10	31	17	9	67

Table 4.2 Exit Questionnaire in English (2017)

EXIT QUESTIONNAIRE: 66 PARTICIPANTS / February 14, 2017						
Number	Questions	Agree	Partially Agree	Disagree	Completely Disagree	Total
1	I was more anxious and nervous compared to before doing the speaking test.	24	15	13	14	66
2	I was informed about the procedure of the test.	55	9	0	2	66
3	I was very comfortable before the test.	20	24	18	4	66
4	I did not have any idea about the assessment scale (rubric).	5	10	18	33	66
5	I was prepared in class (through strategies and activities) for the speaking test.	51	15	0	0	66
6	I had difficulty in expressing myself clearly and in an adequate way during the test.	23	24	17	2	66
7	I was anxious and nervous during the speaking test.	24	18	19	5	66
8	I was encouraged and motivated by the teacher during the test.	21	28	10	7	66
9	The tasks/activities in the test were types I had never encountered before.	5	10	23	28	66
10	I believe thet the speaking activities in classes were insufficient.	2	10	29	25	66
11	After the test, I realized that I needed more speaking practice.	31	29	6	0	66
12	I felt more comfortable doing this new oral exam rather than previous one.	20	29	6	11	66

1) **Advantages:** The advantages that this oral exam proposal holds, according to these learners, can be explained through their answers to Q2, Q3, Q5, and Q8 of these two structured questionnaires.

Q2: I was informed about the procedure of the test.

Figure 4.1 Q2 Entry questionnaire results

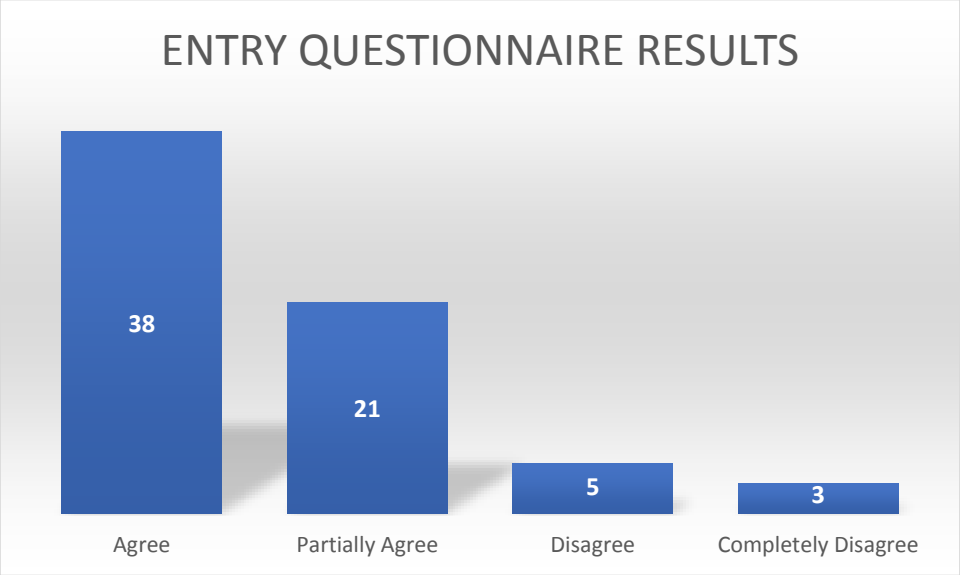
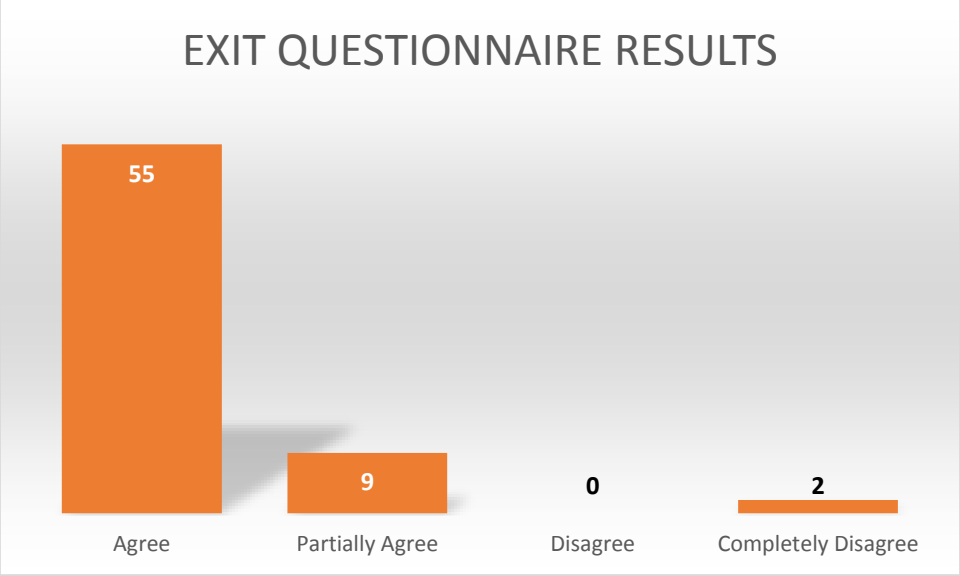


Figure 4.2 Q2 Exit questionnaire results



Q2 served the purpose of discerning whether learners held or did not hold basic, compulsory and necessary knowledge about the procedures that these oral exams entailed. From the entry questionnaire, it can be observed that 88% of this group

showed agreement (agree - 57%; partially agree - 31%) with the content of this inquiry for their first spoken test. After doing the exit questionnaire, which corresponded to this oral exam proposal, this trend of accordance (agree - 83%; partially agree - 14%) not only continued, it increased (i.e. by 9%) with 97% as the final percentage. These results indicate that these students were well-informed about the procedures to be followed during the current oral exam format and the new verbal assessment scheme as well. This fact is advantageous because the learners are well aware of all the complexities present in these exams, which means that time spent on explaining this point to them is never misused. Nonetheless, the percentage of undergraduates who was pretty conscious of the whole process involved in this new verbal assessment scheme was significantly higher (i.e. +26%; total of 83%) than the number of pupils (i.e. 57%) who knew all the procedure behind the current oral exam. Therefore, almost all of these students pretty much knew in advance what specific steps they had to follow during this spoken test and how their performances were going to be graded.

Q3: I was very comfortable before the test.

Figure 4.3 Q3 Entry questionnaire results

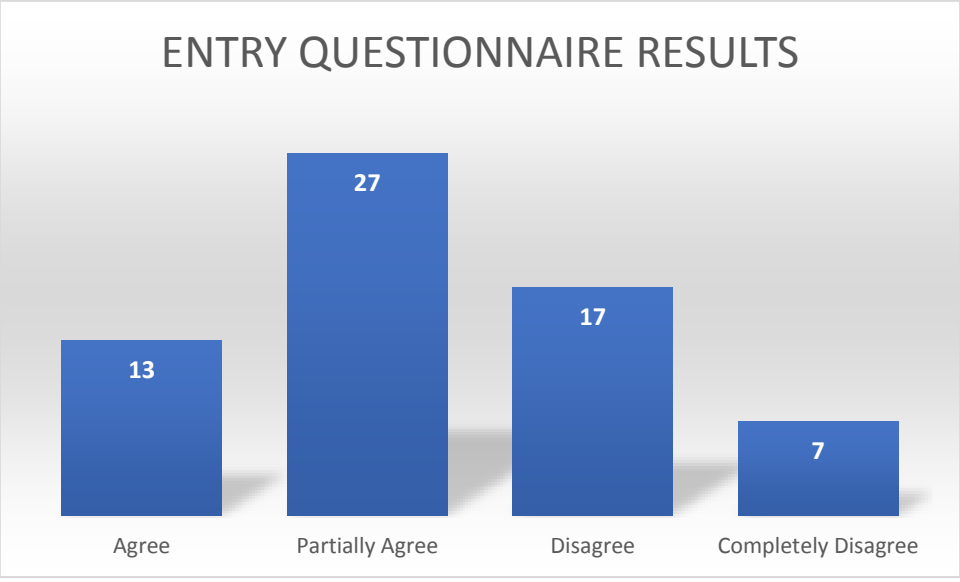
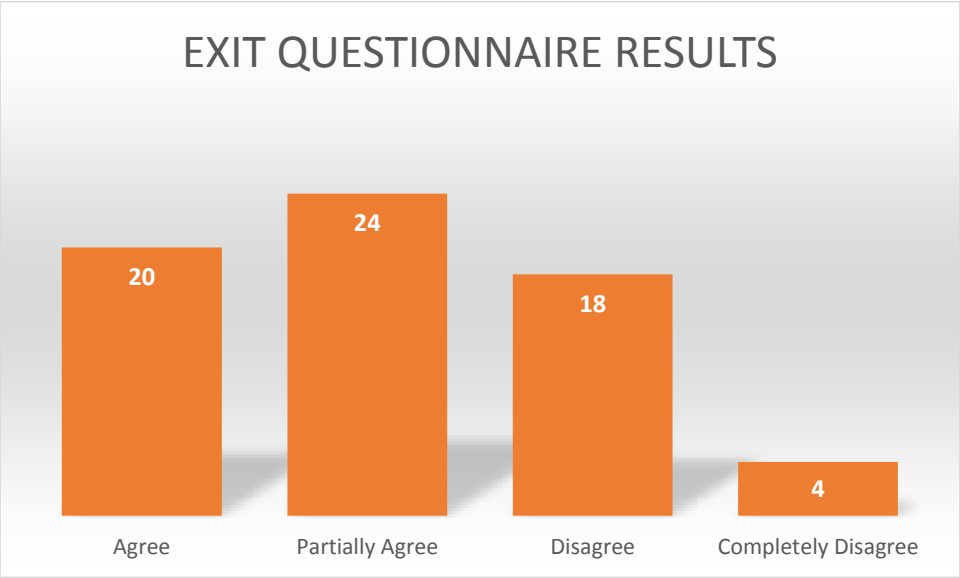


Figure 4.4 Q3 Exit questionnaire results



The level of anxiety of these participants was measured on the three stages that correspond to this spoken test – "before", "while" and "after" doing the verbal examination; Q3 covered the "before" bit and these students phases of calmness during it. In the entry questionnaire, the numbers showed that 62% of these learners accept to some extent (partially agree - 42%; agree - 20%) that they were very comfortable before the test. Likewise, after checking the figures from the exit questionnaire, there was a partial acceptance (partially agree - 37%; agree - 30%) regarding the level of anxiety before doing this spoken test in 67% of this collection of learners. Thus, most of these students felt most at ease before doing the current oral exam as well as this verbal assessment proposal. However, for the exit oral exam, there was an increase in the level of confidence (i.e. +10%; entry exam - 20%, exit exam - 30%) in seven of these learners as they indicated that they agreed with the utterance in the question.

Q5: I was prepared in class (through strategies and activities) for the speaking test.

Figure 4.5 Q5 Entry questionnaire results

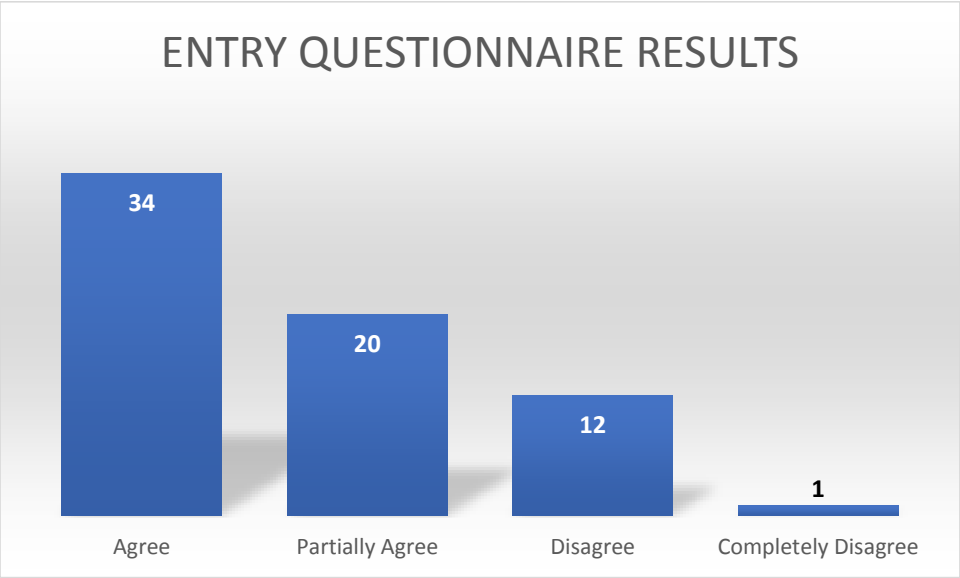
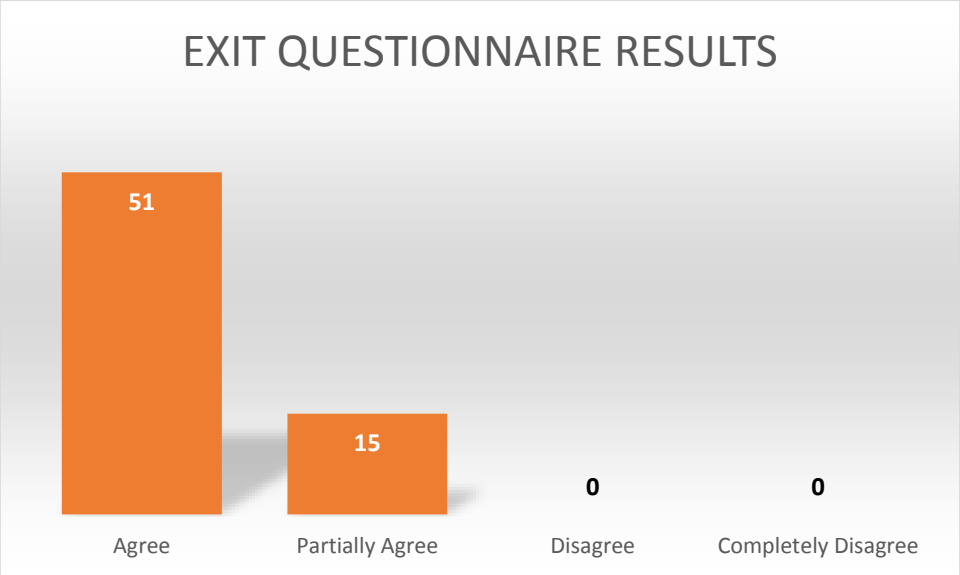


Figure 4.6 Q5 Exit questionnaire results



Receiving appropriate training for this oral examination was a vital component of this research, particularly for the stakeholders involved as the teaching practices of the instructors had to reflect clear priorities for and share common goals with the learning process of these pupils. The main objective of Q5 was grasping the points of view of these participants concerning this issue. In the entry questionnaire, 81% of these undergraduates expressed (agree - 51%; partially agree - 30%) that they had been prepared in class for the existing speaking test. Similarly, a remarkable 100% of these

learners acknowledged (agree - 77%; partially agree - 23%) that they had been exposed to strategies and activities related to the content of this oral exam proposal before doing it. Nevertheless, what truly differentiates this new verbal assessment scheme from the current one is that none of these students pointed toward a lack of appropriate training for this innovative spoken evaluation.

Q8: I was encouraged and motivated by the teacher during the test.

Figure 4.7 Q8 Entry questionnaire results

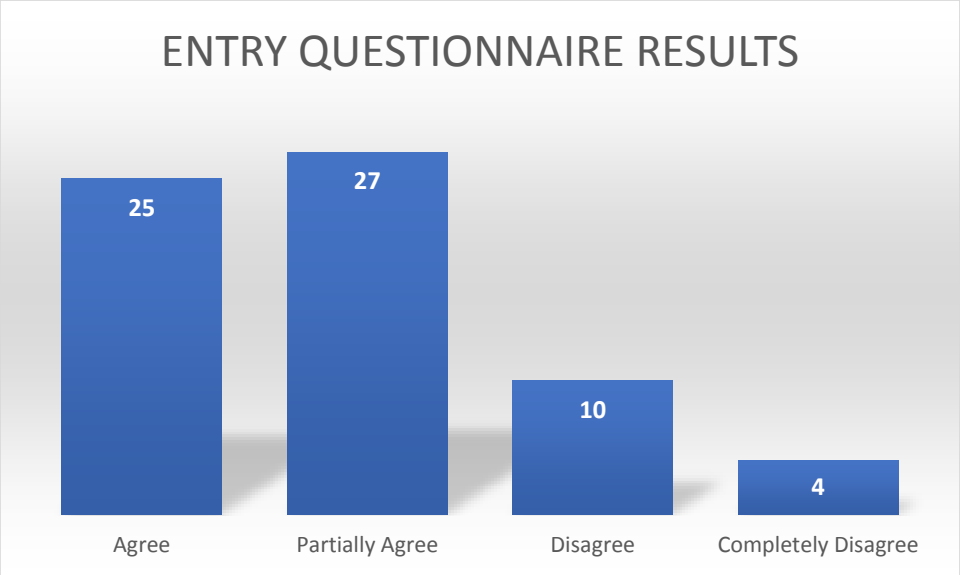
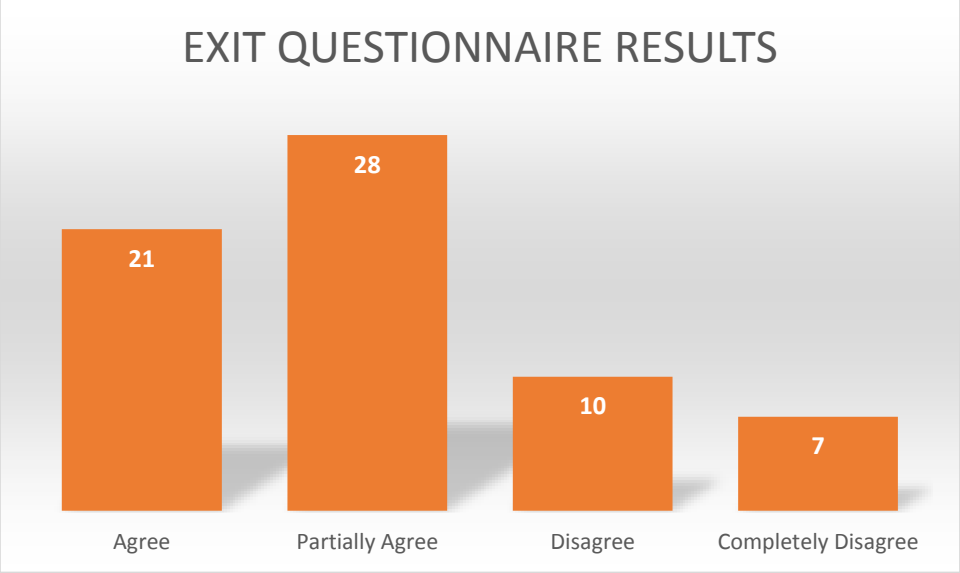


Figure 4.8 Q8 Exit questionnaire results



Q8 was included by this pair of academics to make apposite realizations on how well the academic in charge of this spoken test used the existing format concerning satisfactory yet brief bonding with these learners and their appropriate stimulus. In the entry questionnaire, 79% of these undergraduates was encouraged and motivated by their own teacher (partially agree - 41%; agree - 38%) during their first test (i.e. current oral exam). Alternatively, from the data gathered in the exit questionnaire, only 74% of these pupils considered that the "other" instructor had provided fine and sound handling (partially agree - 42%; agree - 32%) of the new spoken test arrangement. As a result, this mild decrease in the percentage of conformity (i.e. -5%) reflects that the presence of a different teacher during this second examination definitely altered the behavior these students had had with their primary EFL instructor; even so, this one digit figure is well within reasonable and satisfactory boundaries for these social scientists.

2) **Disadvantages**: After doing this oral exam proposal, these students verbalized the problems that they recognized on this oral exam proposal as disadvantages by answering to Q1, Q4, Q6, Q7, and Q9 from this pair of structured questionnaires.

Q1: I was more anxious and nervous compared to before doing the speaking test.

Figure 4.9 Q1 Entry questionnaire results

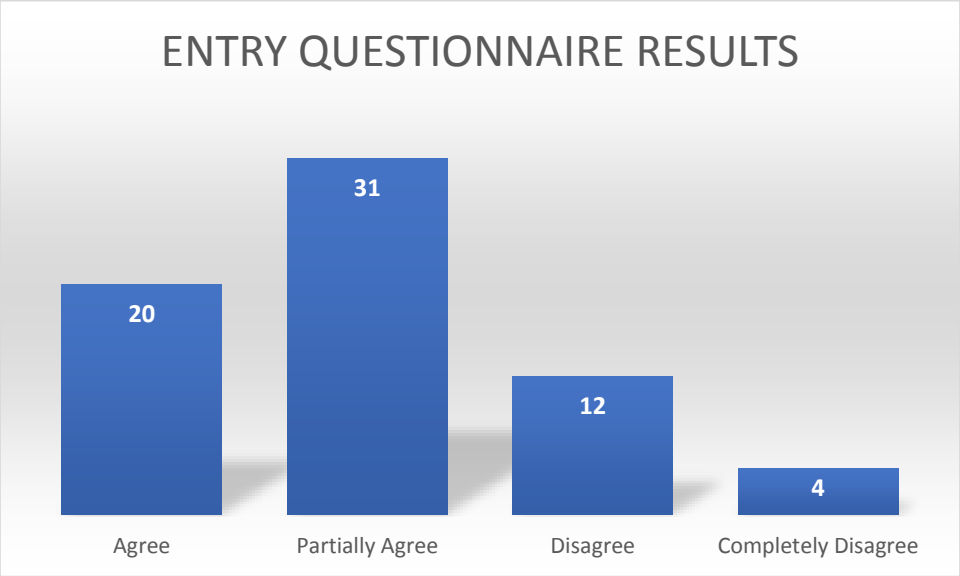
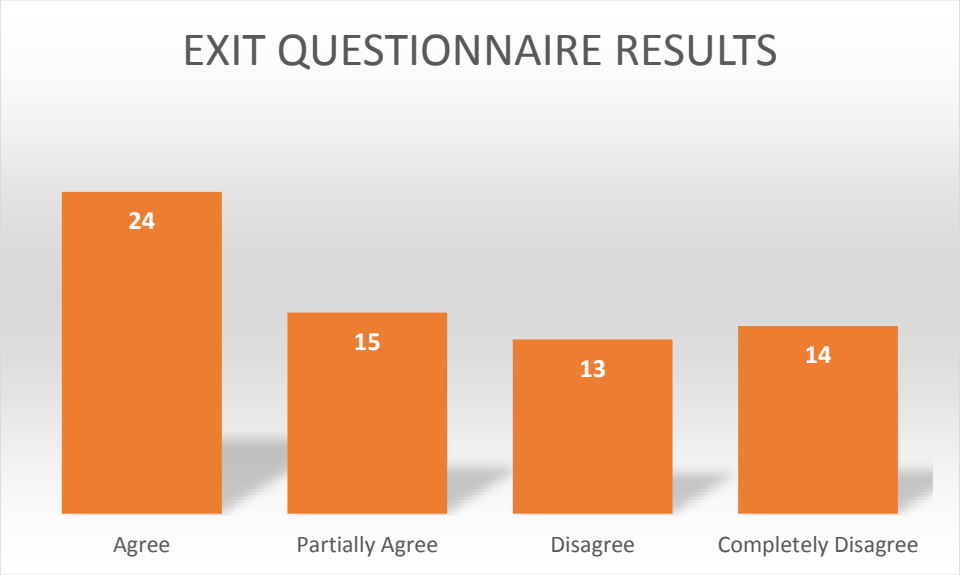


Figure 4.10 Q1 Exit questionnaire results



Q1 was employed to measure the level of anxiety that these participants still bore "after" doing this new verbal examination. In the entry questionnaire, 76% of these learners admitted (partially agree - 46%; agree - 30%) being more anxious and nervous compared to before doing the speaking test. In contrast, the information from the exit questionnaire specified that only 59% of these students felt an increase in their levels of anxiety and nervousness (agree - 36%; partially agree - 23%) after leaving the classroom wherein they did this oral exam proposal. This means that a satisfactory number of these undergraduates gained assurance and serenity for their second vocal examination thanks to the meticulous and detailed training received throughout this study. Nevertheless, this percentage (i.e. 59%; a decrease of 17 percentage points from 76%) stands for a number which is not completely reassuring for these academics as it means that only 39 people – out of 66 – from this group felt confident enough to announce that they had done well in this new spoken test. In order for this numerical fact to be qualified as an advantage of this new format over the current one, these scholars consider that the percentage should have reached at least 75%.

Q4: I did not have any idea about the assessment scale (rubric).

Figure 4.11 Q4 Entry questionnaire results

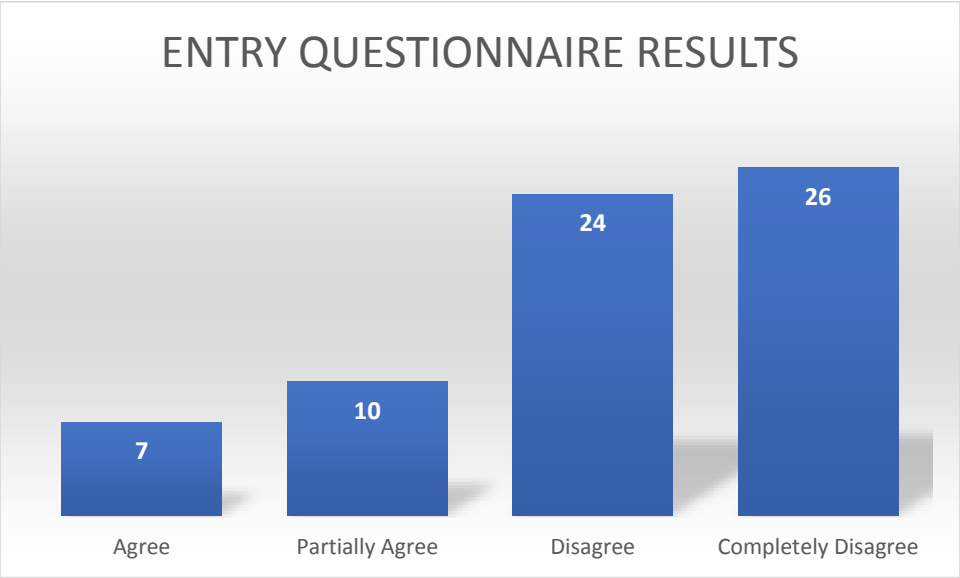
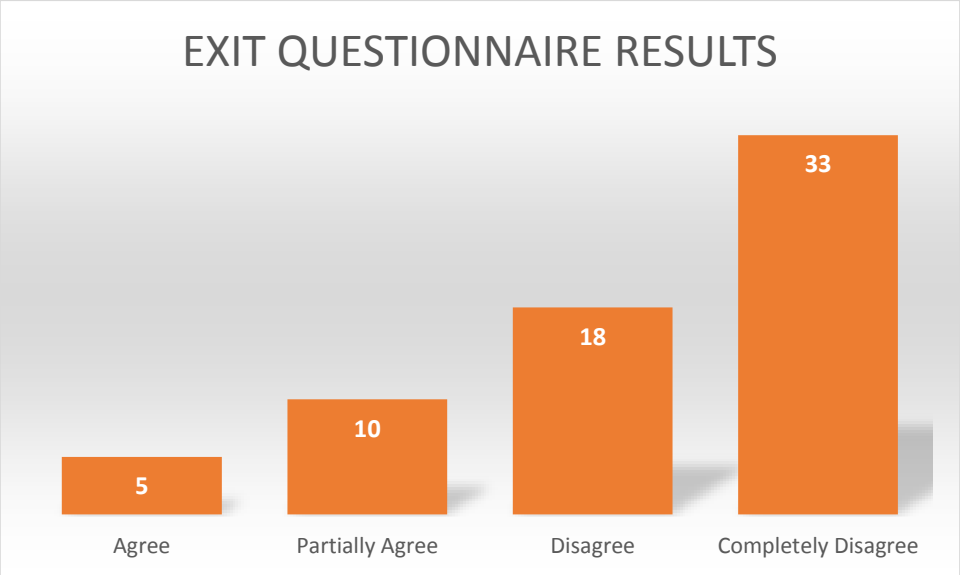


Figure 4.12 Q4 Exit questionnaire results



Q4 asked these learners if they had total knowledge about the rubric with which their performances were going to be graded in both exams. In the entry questionnaire, 75% of these undergraduates mentioned that they knew the assessment scale (completely disagree - 39%; disagree - 36%) used in their first spoken test. In the same way, the exit questionnaire proved that 77% of these pupils affirmed being acquainted with the different components of the rubric (completely disagree - 50%; disagree - 27%) designed for this oral exam proposal. The reader can have a look at this ample,

comprehensive, effective evaluation tool in page 135, under the name of Appendix 6. Even though the assertions made by these students demonstrated that they were quite familiar with the assessment scale of this new verbal evaluation scheme, the analysis done by these researchers in relation to the concrete responses of these learners identified not having adequate understanding or not possessing sufficient comprehension of the grading system as an expected and likely disadvantage in case this oral exam proposal is instilled in the near future.

Q6: I had difficulty in expressing myself clearly and in an adequate way during the test.

Figure 4.13 Q6 Entry questionnaire results

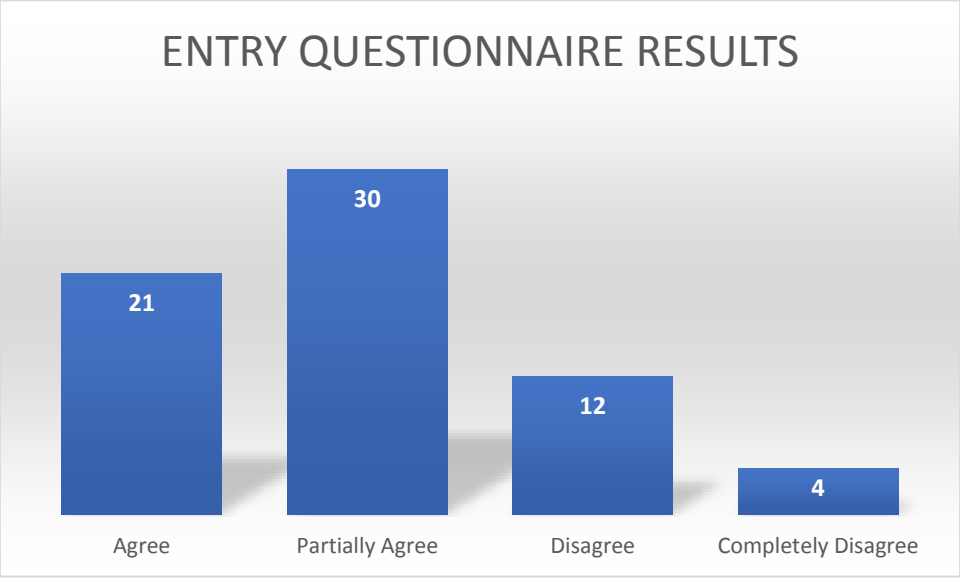
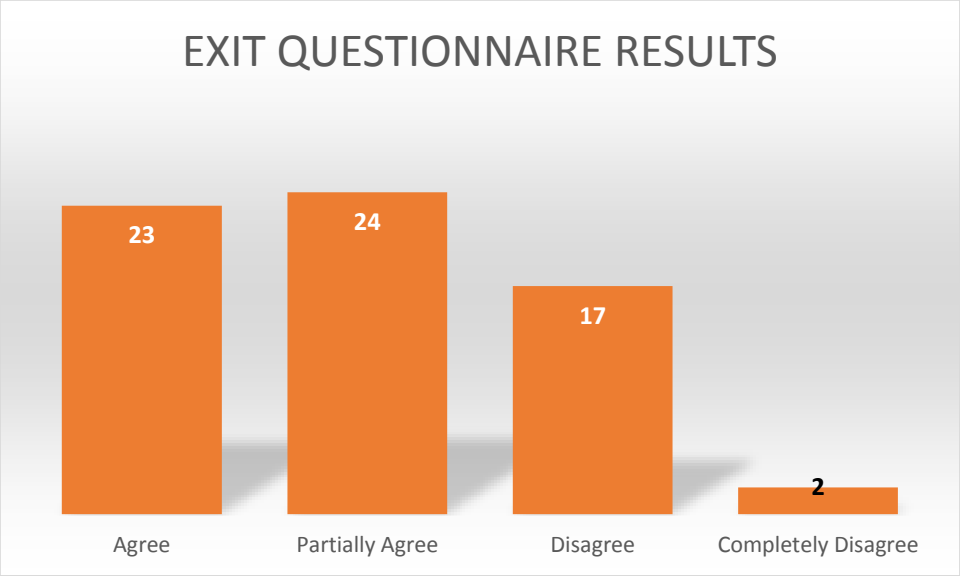


Figure 4.14 Q6 Exit questionnaire results



Q6 was asked to make observations linked to the real linguistic adeptness point (i.e. speaking competence overall) that these learners reached because of this intervention. In the entry questionnaire, 76% of these undergraduates proclaimed (partially agree - 45%; agree - 31%) that they had difficulty in expressing themselves clearly, and in an adequate way, during their first spoken test. Before showing the numbers from the exit questionnaire, it is noteworthy to specify that the results of this second questionnaire and the previous one handed over crucial information which ended up being quite alike. This similarity lied on the detrimental fact that all throughout this oral exam proposal over seventy percent (i.e. 71% pertaining to 36% of partial agreement and 35% of agreement) of these students could not express their ideas well, nor could they have effortless access to the appropriate lexical exponents or the necessary sentence structures to communicate their viewpoints fairly well. Even if there is a 5% decrease – as the absolute amounts for the second spoken test confirmed – in the quantity of people who felt unable to perform acceptably in this exam, these EFL instructors are genuinely concerned about this final figure (i.e. 71%) since it tags this characteristic as a truly negative trait from this new verbal assessment scheme that could only be resolved via continuous and effective in-class training.

Q7: I was anxious and nervous during the speaking test.

Figure 4.15 Q7 Entry questionnaire results

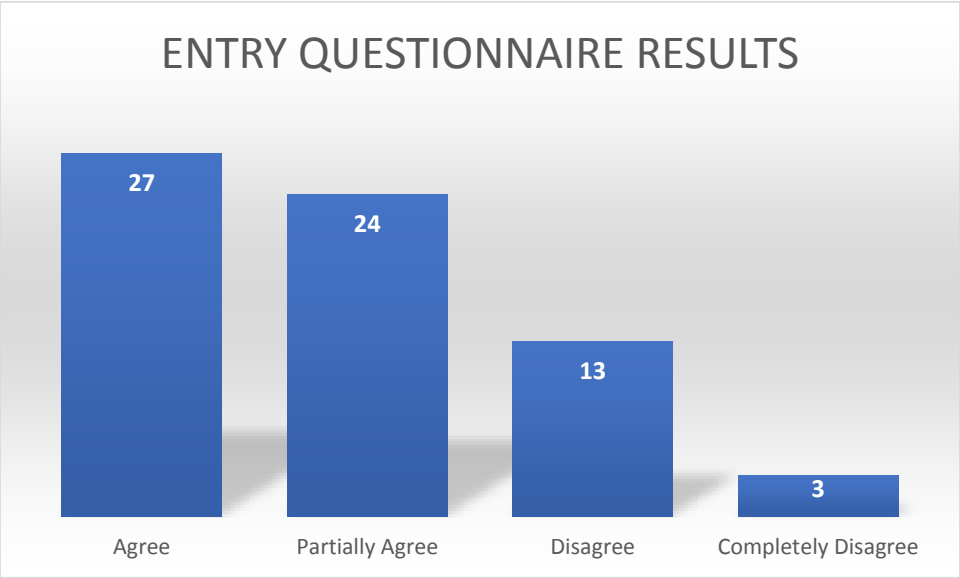
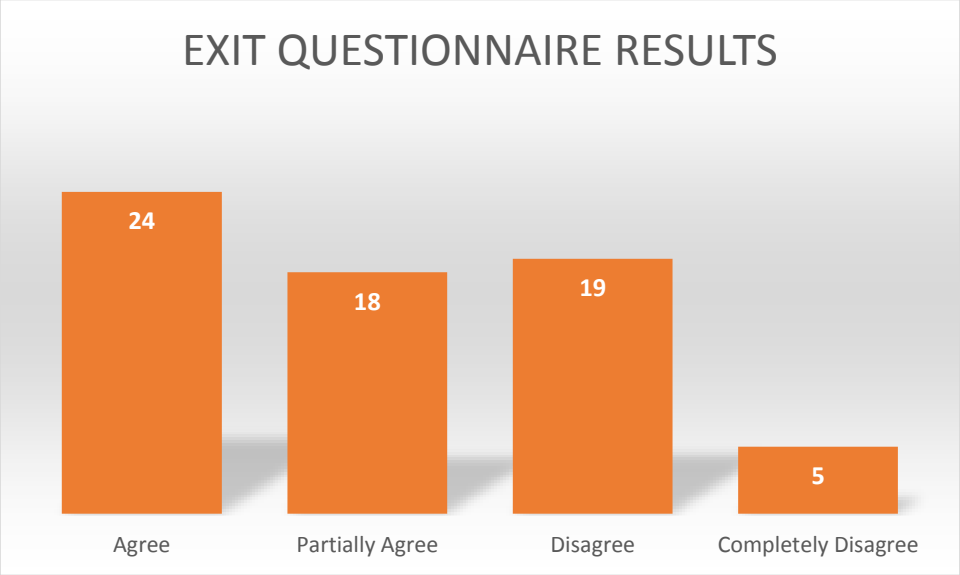


Figure 4.16 Q7 Exit questionnaire results



These researchers remind the reader that level of anxiety of these participants was measured "before", "while" and "after" doing the verbal examination; Q7 helped these academics with the estimation of the "while" part. In this query from the entry questionnaire, 76% of these undergraduates concurred (agree - 40%; partially agree - 36%) with the proclamation of being anxious and nervous during their first speaking test. Meanwhile, for the exit questionnaire, 63% of these learners consented (agree - 36%; partially agree - 27%) with experiencing tenseness and feeling unease while performing on this oral exam proposal. Even though the percentage decreased by 13% (i.e. from 76% to 63%) on this question, it gives the reader an evident idea about how the minds of these students functioned throughout this second spoken assessment – with preponderant levels of anxiety and nervousness. Therefore, being indisputably precise and complicated enough to make pupils feel nervous and anxious are indeed negative traits – for these students, not their teachers – that this new oral evaluation possesses.

Q9: The tasks/activities in the test were types I had never encountered before.

Figure 4.17 Q9 Entry questionnaire results

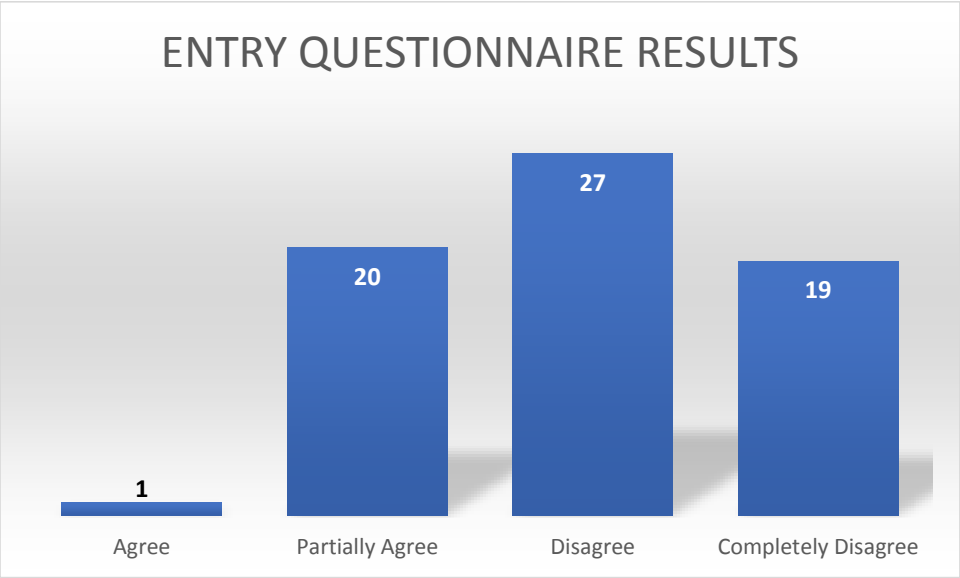
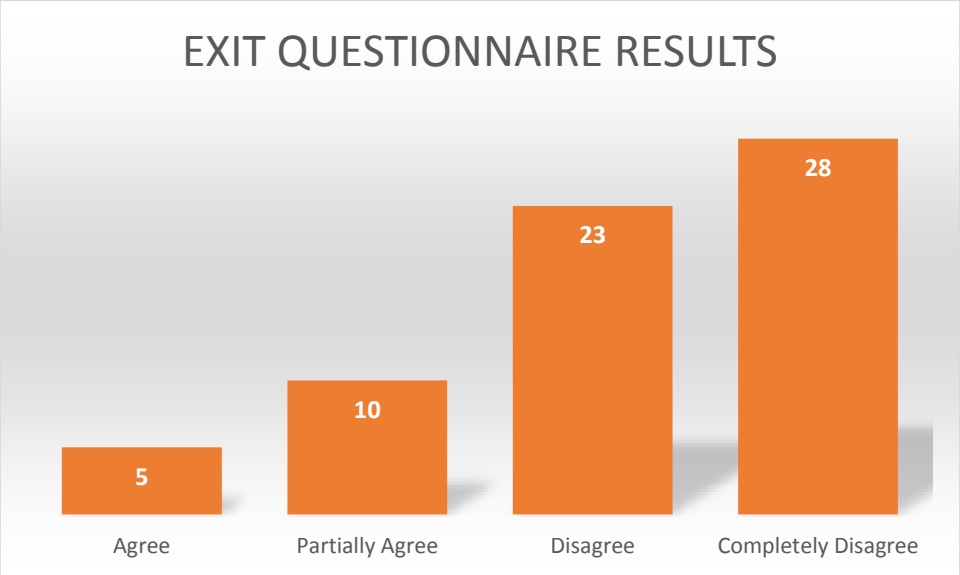


Figure 4.18 Q9 Exit questionnaire results



For quantifying the quality and thoroughness of the training received during class sessions, these academics resorted to Q9 whose wording is quite different from Q5, but its underlying rationale remains equal. The results of the entry questionnaire for Q9 stated that 68% of these students had already encountered and worked with the activities and the tasks (disagree - 40%; completely disagree - 28%) that were present in and carried on during their first spoken test. In the same way, according to the exit questionnaire, 77% of these learners drew attention to having been frequently exposed

to the tasks and activities (completely disagree - 42%; disagree - 35%) available on this oral exam proposal. In fact, these researchers trust that these pupils became rather familiar with those classes of exercises right through their coaching and tutoring sessions.

The fact that there was an increase of 9 percentage points (i.e. from 68% to 77%) from the entry questionnaire to the exit one signifies an attractive improvement in this regard, which definitely wound up facilitating these undergraduates performances during this new verbal assessment scheme. However, if the tasks and activities of this or any other future examination are never practiced nor taught during class, then those undergraduates taking that course evaluation will surely fail. This staggering reality, which can only be altered by pensive EFL instructors, might be pronounced as a fundamentally decisive and significant flaw in the design for this oral exam proposal.

3) **Recommendations:** The recommendations that, according to these learners, can help turn this new verbal assessment scheme into an even more effective grading tool were found in the answers to Q10 and Q11 from these structured questionnaires.

Q10: I believe that the speaking activities in classes were insufficient.

Figure 4.19 Q10 Entry questionnaire results

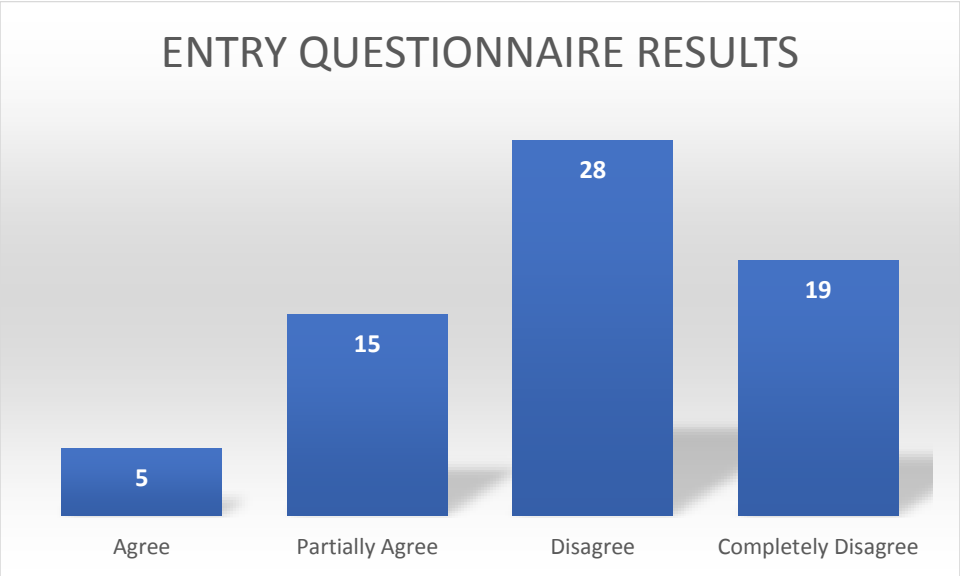
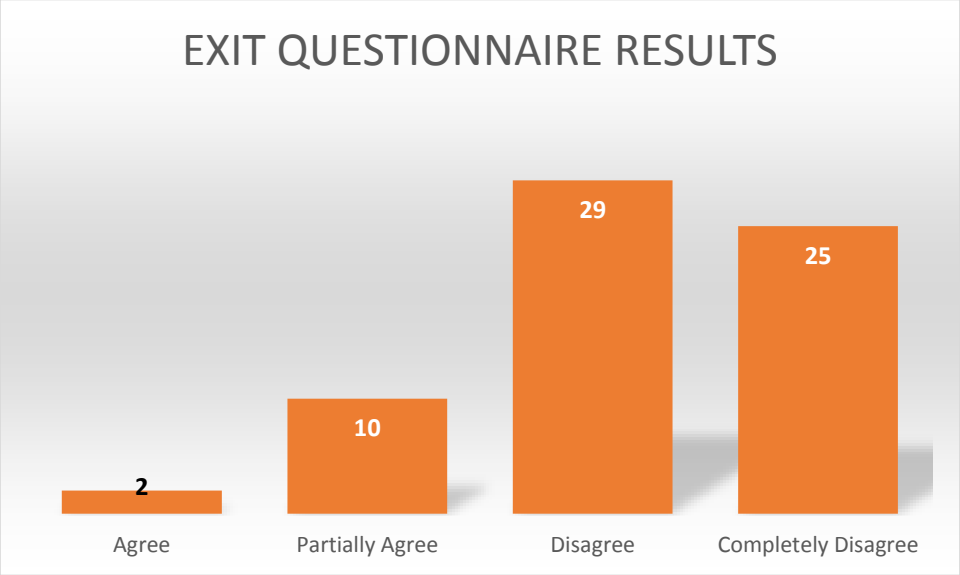


Figure 4.20 Q10 Exit questionnaire results



These researchers brought Q10 into play to evaluate the sufficiency or insufficiency, in terms of quantity, of the verbal activities conducted during these classes with the intent of preparing these learners for their first spoken test as well as for the proposed oral exam. From the data gathered in the entry questionnaire, 70% of this sample indicated that the number of speaking activities done in classes were enough (disagree - 42%; completely disagree - 28%) to do well in the current oral exam. Likewise, after revising the information collected from the exit questionnaire, 82% of these learners asserted that they had been exposed to an adequate amount of examples (disagree - 44%; completely disagree - 38%) on the subject of the speaking activities available on this oral exam proposal. This 12% boost in the percentage points (i.e. from 70% to 82%) establishes that the number of verbal tasks handled and managed during the coaching and tutoring sessions, carried out throughout this study, were sufficient so that this collection could proceed fairly well in this new verbal assessment scheme. Hence, using an appropriate number of speaking activities in classes is something that the EFL instructors who will be applying this oral exam proposal in the near future should really take into consideration.

Q11: After the test, I realized that I needed more speaking practice.

Figure 4.21 Q11 Entry questionnaire results

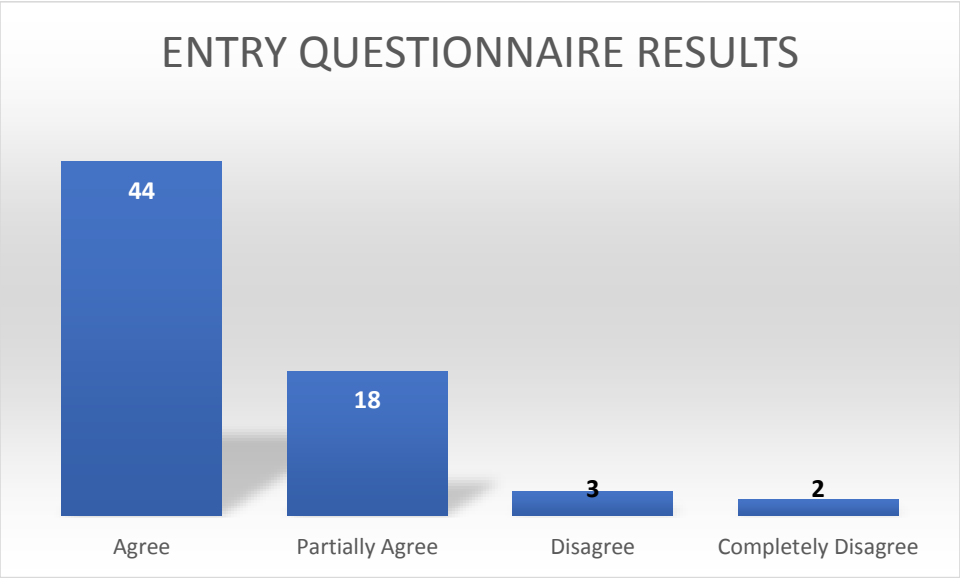
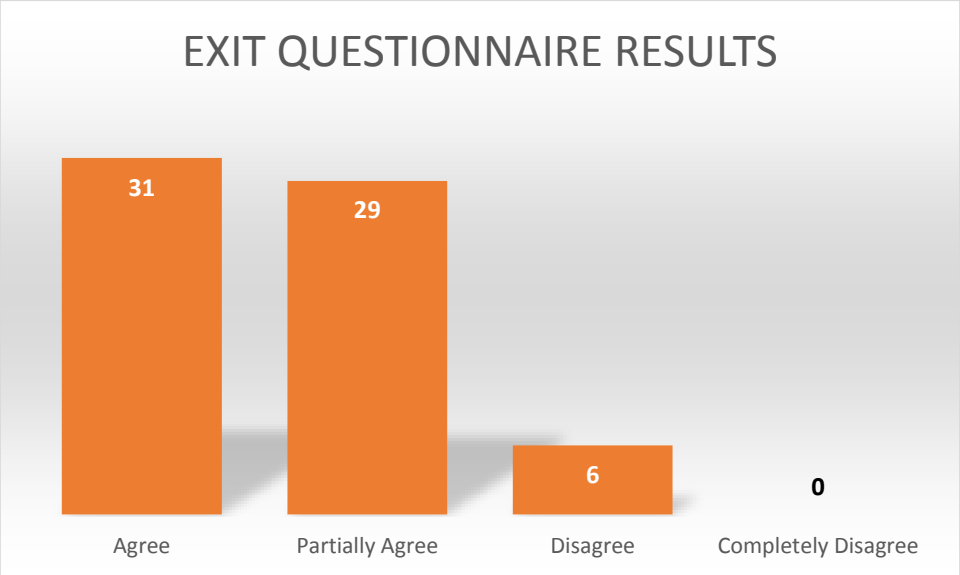


Figure 4.22 Q11 Exit questionnaire results



To find out the actual proficiency level that these learners had attained due to the intervention they had been part of, these researchers included Q11 in these structured questionnaires. From the entry questionnaire, 93% of these learners declared (agree - 66%; partially agree - 27%) that after doing the first spoken test, they realized they needed more speaking practice during their classes. Alternatively, after analyzing the exit questionnaire, 91% of these undergraduates avowed (agree - 47%; partially agree - 44%) that they felt unconfident about having conceived and produced enough language

for this oral exam proposal. As a consequence of the data projected by these two numbers (i.e. 93% and 91%), it can be unequivocally affirmed that these partakers recognized an actual need of more speaking practice for both assessments. In an attempt to improve in that respect, these instructors should have utilized more tasks – during their classes in general and for their guidance periods especially – that resembled the likes of the ones this particular crowd was going to encounter all through these oral exams. Therefore, it can be inferred that a great number of these students has suggested working more on speaking activities – mainly in class – to improve the proficiency level and the fluency stage that they could wield during any oral exam. Additionally, despite the format of the examination, students recommended designing new and innovative activities that can contribute – to a great extent – to enhance their performances in this standardized oral exam or any other spoken test.

4) **Approval:** This group of students approved or disapproved the use of the oral exams applied during this study by means of responding to Q12 of these structured questionnaires.

Q12: I felt more comfortable doing this new oral exam rather than previous ones.

Figure 4.23 Q12 Entry questionnaire results

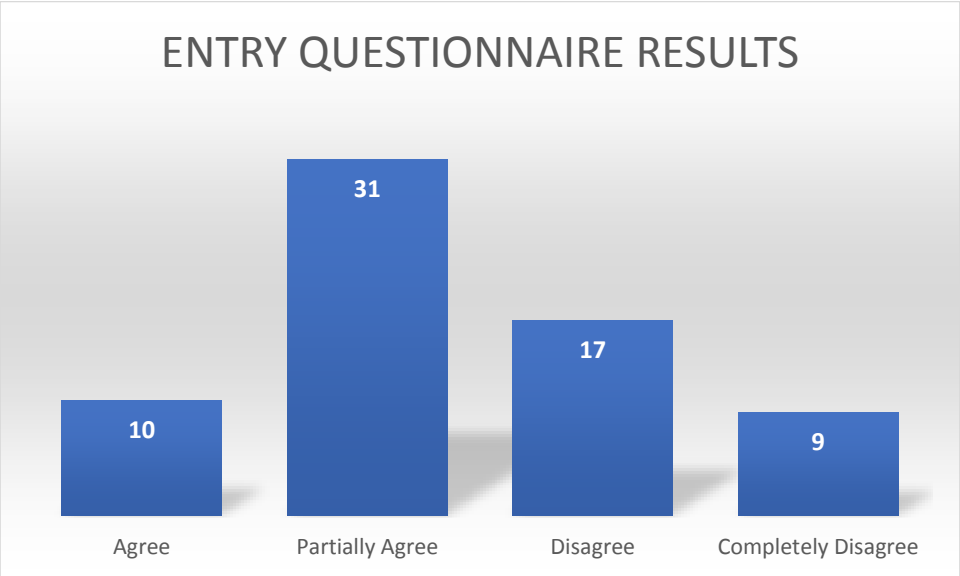
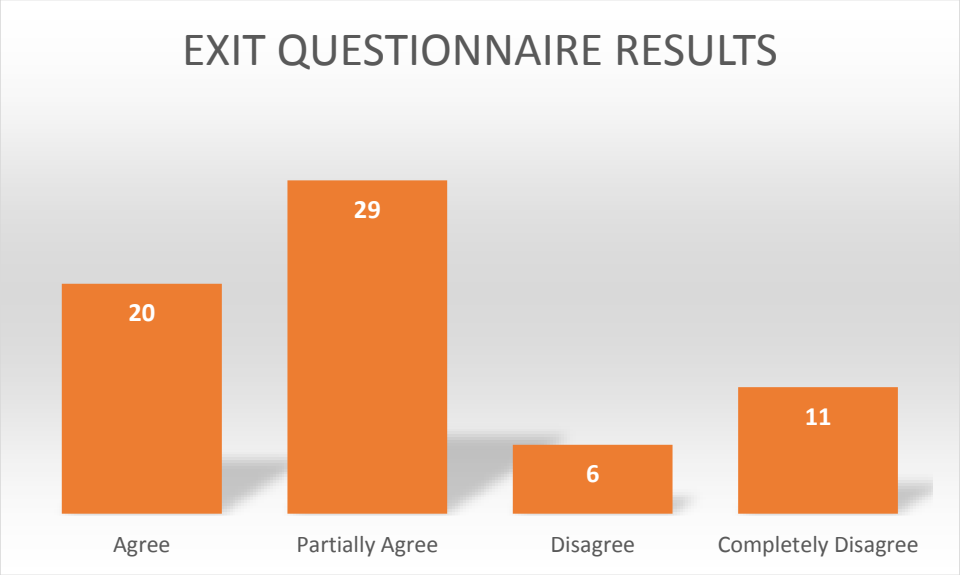


Figure 4.24 Q12 Exit questionnaire results



Q12 allowed this cluster of learners to affirm or deny that they felt more relaxed with these formats instead of with any other one they had been tested upon and evaluated with. From the entry questionnaire, it is valuable to mention that the percentage of participants who were in agreement (partially agree - 46%; agree - 15%) with this comment of feeling more relaxed with the present oral exam arrangement than with any other format adds to 61% hence the performance of these undergraduates should have reflected that fact positively, in terms of grades, because they were at ease with being graded through this existing verbal assessment design. After considering the answers available on the exit questionnaire, it must be averred that 74% of these pupils concurred with being more comfortable using this standardized format of an oral exam (partially agree - 44%; agree - 30%) than any other previous design, including the current layout. For that reason, the final perception of these students towards this oral exam proposal is of the positive type therefore they approve and support its format, its length, its process, and its grading methodology.

4.7 Interpretation of data from questionnaires

Here is a list that sums up the most important deductions reached after analyzing the figures provided by the entry and exit questionnaires. It is necessary and relevant to specify that these interpretations only consider and are based on the highest percentages

revised. In spite of that fact, these next statements make especial focus on the responses directly linked to the data gathered from the exit questionnaire.

1) For students to do well in this oral exam proposal, they are required to hold an adequate knowledge regarding the procedure involved in this new verbal assessment scheme, and they must also try to stay calm before taking part in this innovative spoken test. Additionally, so their learners can perform well in this oral exam proposal, the EFL instructors in charge of this vocal examination are obliged to apply effective and efficient speaking development strategies as well as use a sufficient amount of oral activities related to the content of this new verbal assessment before doing it. Finally, the scenario wherein another educator, not the one who has been teaching them, does the grading of their spoken test is something that these students have to get used to since any serious international exam will place them in front of an EFL instructor with whom they have no acquaintance at all.

2) The learners will not have a good performance on this oral exam proposal if they do not feel confident enough about their speaking skill to have a good grade after doing it; in other words, if they still feel anxious and nervous after taking part on this innovative spoken test, then they will definitely not have a score with which they would agree. The number of undergraduates who will do badly in this new spoken test will also increase considerably if they are not aware of how the rubric and its components will be used to grade their vocal performances. Furthermore, the learners whose overall speaking competence has not been properly developed inside the classroom are certainly going to have difficulty to express their ideas well since their lexis, grammar, or both will not allow them to communicate their perspectives as fluently as they would like to. As a final point, feeling anxious and nervous while doing this oral exam proposal is not a desirable condition or recommendable state for those pupils whose final grades will depend on their effective and efficient completion of this new spoken test hence their EFL instructors must regularly expound them, especially in their class sittings, to the tasks and activities available on this new verbal assessment scheme.

3) If the EFL instructors who will apply this oral exam proposal want their students to have the best grade possible, they ought to plan well in advance the number – as well as the attributes – of the verbal activities that their pupils will participate on during their regular classes as this point is vital and critical for their actual success while handling the two components of this innovative spoken test. Likewise, these teachers should also include and use more speaking development strategies that are not only

adequate, but adept for the elements of this new verbal assessment scheme in their lesson plans.

4) This oral exam proposal and all its complexities were approved by almost three quarters (i.e. 74%) of the whole sample. Therefore, these researchers can claim without a shred of doubt that this oral exam proposal was an utter success.

Considering all the deductions explained above, it can be concluded that the only time when these students were at ease was before doing this oral exam proposal as they were uncomfortable and stressed all throughout this spoken test; even after it, in a good number of cases. Consequently, it is quite apparent that having anxiety and / or being nervous during an oral exam has nothing to do with being properly trained, feeling well-prepared and knowing everything there is to know (i.e. format and rubric) about a verbal assessment, at least for this particular group. It is also quite important to state that one thing is to be very comfortable with doing something (i.e. having a high level of confidence to carry an action out) and another, very different thing is not to be anxious and nervous about it (i.e. getting used to performing on a deed or starting to gain self-assurance while acting upon it), as these researchers were able to notice in the behaviour of some of the participants from this sample whilst doing this oral exam proposal.

On the most negative side of these assumptions, it has to be mentioned that the training provided by these teachers regarding strategies for speaking development proved to be insufficient as almost all the participants reported not having emitted enough exponents of the target language during this oral exam proposal, which means that they would have definitely benefited of having practiced various more speaking tasks in class.

Finally, these researchers have confidence in the fact that these learners should have done moderately well – academically speaking – in this oral exam proposal because they declared feeling comfortable with this final arrangement. However, the only possible way to prove that statement is correct will be checking and rendering their grades for this particular progress achievement test.

4.8 Analysis and interpretation of data from numerical results

The data collected in this study (i.e. before and after the intervention) provided these researchers with the quantitative results that the reader will examine promptly in order to continue building proper and complete understanding of the efforts embraced

by this educational project. The aforementioned information will be exhibited on this dissertation after completing the following statistical procedure:

1. The process matrix is constructed respecting the data obtained in the research.
2. The values of the following statistical measures are calculated: arithmetic means, standard deviation, and standard error for the media.
3. The value of the statistical significant test known as student's t-test (*t*) is calculated.
4. The analysis of the value of *t* is carried out.
5. A decision is made based upon the interpretation of *t*, which basically means that the result obtained is contrasted against the null hypothesis for its acceptance or rejection.

To fully explain the quantitative data produced during this educational project, it is important to describe and understand the nature of the information provided by this study. Hence, these social scientists will from now on be regarded as "RESEARCHER A" (i.e. Critical Thinking Expert) and "RESEARCHER B" (i.e. Experienced Oral Examiner for International Exams); this declaration will turn the presentation of their groups results into a much more understandable and forthright task. It is also vital to notify the reader that these academics resolved unanimously to only take into account the scores produced by sixty-five (65) students from this sample. This decision suggested itself when two of the students from GROUP "B" did this oral exam proposal a day after their classmates from both groups had done it. Therefore, even though their opinions regarding the new verbal assessment scheme were valuable and valid, their scores feedback could not be considered the same way as these learners might have been informed on – some or all of – the different questions that are part of this oral exam proposal, which is definitely an unfair and undeniable advantage over their classmates.

It is now time to proceed with the statistical significance test chosen by these researchers for this study. These academics are letting the reader know that the results collected from the entry exams and the exit exams of "RESEARCHER A" and "RESEARCHER B" were separated in two different tables. This decision was made by these academics in order to explain the data collected for this study without any bias and – most importantly – in an accurate way.

In the specific case of "GROUP A", the results from the entry evaluations to the exit examinations varied positively from lower numbers to higher figures on the whole.

This change means, in general terms, that these students performed better on this oral exam proposal than on the current format for the speaking test used by this language center.

The *t* value calculated is *-10,52*. (APPENDIX 12) This value is well outside the boundaries from the region of *not acceptance* after comparing it to the critical limits, which according to the 0.05 significance level or 0.975 critical value for 35 participants are any values that are greater than ± 2,030 in the Student's *t* Distribution Table. (NIST/SEMATECH, 2013) This numerical outcome clearly reflects the complete acceptance of the new verbal assessment scheme by these students.

The end results these learners got in their exit oral exam (i.e. oral exam proposal) were much better than the ones obtained in their entry spoken test thus the null hypothesis for this set of undergraduates was rejected. Therefore, "RESEARCHER A" could affirm that his students fully approved this new verbal assessment scheme as well as the strategies used throughout this intervention because his pupils showed competent domain of them in the process of this vocal examination.

Let's continue with the scrutiny of the information gathered from the group that was supervised by "RESEARCHER B". The results of the entry and exit examinations for the constituents of "GROUP B" differed from higher to lower quantities and that is the main reason why these researchers decided to examine these figures separately from those numbers belonging to "GROUP A". If this information set (i.e. "GROUP B") had been analyzed together with the figures from the other assembly (i.e. "GROUP A") then the general statistics of the study would have been evidently affected, and what is worst in an extremely negative way.

The *t* value calculated is *7,91* (APPENDIX 13) because the end results gotten by these pupils in their exit oral exam (i.e. new verbal assessment scheme) were different or less – as a general rule – than the first scores obtained in their entry oral exam. This value once again is well outside the boundaries from the region of *not acceptance* after comparing it to the critical limits, which according to the 0.05 significance level or 0.975 critical value for 30 participants are any values that are greater than ± 2,042 in the Student's *t* Distribution Table. (NIST/SEMATECH, 2013) However, this statistical product evidently mirrors that the students of this specific collection had stern problems during the process of this oral exam proposal, especially in the "interactive turn".

The scores obtained by these students after doing this new verbal assessment scheme were dissimilar and lower compared to the marks of the current oral exam

consequently the null hypothesis for this set of undergraduates was rejected as well. Except this rejection does not signify that there was a significantly positive change in the sample; quite the opposite, it means that these students had serious difficulties during this innovative spoken test hence the strategies exercised for the EFL instructor in charge of their guidance during this educational project were not quite suitable for the linguistics needs of these particular learners.

To sum up, the quantitative data demonstrated that the students who were coached by the Critical Thinking Expert did much better on this oral exam proposal than the learners who worked under the supervision of the Experienced Oral Examiner for International Exams. This noteworthy occurrence was the result of receiving the whole training as these researchers had originally intended for all of these students and their underlying profit. For that sole reason, the undergraduates in "GROUP A" had the advantage of knowing all the techniques that were necessary to do quite well in this new verbal assessment scheme, and they definitely benefited from it. As the members from "GROUP B" did not follow the same instructional plan, due to their actual English proficiency level which was the main responsible for the negative washback effect that took over them, their end results proved that the initial design had to be completed so that these pupils grades would have increased and shown that they were thoroughly prepared for this oral exam proposal.

To conclude this section of the dissertation, the quantitative data showed that even though a good number of these students (i.e. from "GROUP B" specifically) did not have a good performance on this innovative spoken test, they did not contemplate stating a negative opinion regarding this new verbal assessment scheme. As a matter of fact, these learners never shared a pessimistic remark about this oral exam proposal with these EFL instructors.

CHAPTER 5

5. SUMMARY, FINDINGS, CONCLUSIONS, RECOMMENDATIONS AND PLAN OF ACTION

This chapter reports on different issues related to this educational project, such as a summary of the chain of events that occurred during it, the findings derived from it, the conclusions reached by the social scientists who carried it out, the recommendations that these academics make to those EFL instructors who would have to use it in the future, and the plan of action that these social scientists will attend to in the near future.

5.1 Summary

In the last quarter of 2016, these researchers (i.e. a Critical Thinking Expert and an Experienced Oral Examiner for International Exams) identified the latent, impending need present in the Center for Foreign Languages of the Ecuadorian state university where they worked – Closing its standardized evaluation cycle by including an oral component of the same brand. As such change will undoubtedly affect the learners of this educational establishment, these social scientists decided to gather the perceptions from a particular collection of Pre-Intermediate, B1 students attending classes there apropos this new verbal assessment in order to provide relevant qualitative information to the decision-making authorities of this language center. Additionally, to turn this dissertation into a reliable source worthy of future citing, these EFL instructors applied a statistical analysis to the set of quantitative data collected after measuring the actual performances of these pupils on a pre-test (i.e. current oral exam format) and a post-test (i.e. suggested spoken test scheme) for comparing their previous and current proficiency level in the speaking skill. Those scores were obtained from applying a rubric designed for the sole purpose of grading the criteria upon which this vocal examination proposal is based. These language teaching scholars also refreshed their perspective and updated their appraisal expertise subsequent to being exposed to practical training for exerting the rubrics abovementioned. Not only were both the positive and negative outcomes of this didactic endeavor shared with the reader on this document, its foundations, its theoretical framework, its limitations, its drawbacks, and certain recommendations pertinent to this final instructional task were shown as well.

5.2 Findings

The main objectives for doing this educational project were determining and describing the impact and perceptions of a group of Pre-Intermediate EFL learners (i.e. Ecuadorians, state university undergraduates, attending classes at a Foreign Language Center) on the experience of having their speaking skill properly and accurately evaluated with an appropriate oral exam that is not only concise and aptly organized, but also promotes dealing with genuine exchange of ideas settings by bringing critical thinking skills into play. These two objectives were effectively achieved, fulfilling the expectations of the researchers entirely; though the way these complementary aims were carried out was completely different from what had been originally planned by these EFL instructors.

The first step in this study was to evaluate this set of pupils via an oral exam that followed the regular process of mid-term evaluations from this educational establishment. Such process entailed that these undergraduates were evaluated with the current form of the spoken test and its corresponding rubric, which have been developed by and pertain to this Language Center. From that initial stage, the results showed that the group of learners instructed by the Experienced Oral Examiner for International Exams (GROUP "B" and RESEARCHER "B" correspondingly) felt better and scored better in this entry oral exam. In contrast, the cluster of students that received training with the Expert in Critical Thinking (GROUP "A" and RESEARCHER "A" respectively) did not feel comfortable and scored less in that same evaluation. The social scientists in charge of this study affirm that, apart from the actual and varying level of proficiency in oral competence these learners demonstrated, another core source of these contrasting outcomes is that RESEARCHER "B" had more expertise in the entire process of the original exam – expressly in the use of its rubric – presently employed by this Language Center than the one RESEARCHER "A" possessed in this matter.

The second step required these two aspects from these EFL instructors: 1) to create a different format from the one that is instilled, and 2) to design an analytic rubric which had to be presented to three connoisseurs of EFL teaching and learning related areas to get their qualified feedback. The first factor has already been meticulously described in previous chapters. For the second facet, these teachers repaired to an expert in Curriculum Design, a specialist in Assessment, and a professional in Education; these

didactic authorities approved the design of this rubric and its use. This scale of evaluation was then judged by two other language instructors (i.e. applied to their students) to observe if it was useful, and their responses as well as their comments were also affirmative and constructive.

The third step demanded these teachers to inform their students on the assessment tool they were going to use to grade their performances. Therefore, this rubric was very well explained to these undergraduates in class, so that they can know in advance what specific characteristics were going to be deemed as of good quality in this examination. Furthermore, these pupils were exposed to this rubric – in class again – for their complete understanding and to assess their individual progresses in their verbal communication dexterities. Finally, this rubric was a helpful instructive device for both EFL instructors because it was a sensitive part in the set of tasks for using Critical Thinking techniques that RESEARCHER "A" employed in his teaching hours and it was also present in the classroom sittings of RESEARCHER "B" wherein speaking competence activities were worked on. It is compulsory to state that those tasks, techniques and activities improved the speaking skill proficiency level of some of these learners.

The fourth step was the actual application of this verbal assessment suggestion as the exit oral exam, along with its results analysis. As it was extensively seen on Chapter 4, there is a numerical difference in the results for the entry and exit oral exams of these undergraduates. Learners coached by RESEARCHER "B" showed a variation from higher to lower, which means they performed better with the current format than with this proposal. Conversely, those pupils tutored by RESEARCHER "A" achieved a lower to higher differentiation thereby they did much better on this proposal than on the original arrangement. Even though this divergence is obvious, these researchers need to expose that there was significant improvement in the case of most of these undergraduates, particularly from GROUP "A". In fact, those specific students did not only improve their scores from the entry to the exit oral exam, their less skilled learners attained better grades with this verbal assessment suggestion too. However, some of the skilled pupils from this roster also got finer scores while some others maintained almost the same grade, just showing a little improvement.

The fifth step was reporting on the general findings that this qualitative research has provided these researchers with. Firstly, there is enough empirical evidence for the authorities of this Language Center to consider applying other methodologies, such as

Critical Thinking, in the teaching process of this educational establishment. Moreover, that same data should make them realize that it is time to implement a new, standardized oral exam which should take into consideration the steps proposed in this study. Secondly, in the case of teachers, these fine points present vast information to pursue a new system for their lessons planning wherein Critical Thinking activities can be used to create meaning focus endeavors more than language focus attempts.

Other important findings from this study are the reasons why these researchers envision and assert that GROUP "A" outperformed GROUP "B" within the time that the verbal examination scheme suggested by these investigators lasted for each pair of these learners. Those explanations have been categorised into "coherence", "cohesion", "register", "discourse management" and "interaction" for easy recollection on behalf of the reader.

– What is "**coherence**" for these EFL teachers? It is "when [learners] organize their ideas in a logical (or coherent) way." (Harmer, 2007, p.270)

1) **Coherence:** After contrasting the ideas proposed by the members of GROUP "A" to the ones offered by the constituents of GROUP "B" during this verbal assessment, it was distinctive as well as noticeable that the reactions and impressions of the former were intertwined more properly. In addition, their opinions were more concise, and these viewpoints complemented one another in a sounder way too.

– What do these researchers identify as "**cohesion**"? They identify it as "how things stick together in [conversations]. This can be achieved through devices such as … lexical or grammatical cohesion." (Harmer, 2007, p.270)

2) **Cohesion:** The notions and concepts expressed by the undergraduates of GROUP "A" were clearer and better defined when compared to those shared by their university peers of GROUP "B".

– How do these language instructors define "**register**"? It is defined as "the choice of words in a text or conversation on the basis of topic or tone." (Harmer, 2007, p.281)

3) **Register:** The register the undergraduates of GROUP "A" upheld throughout this spoken test was wider than the one possessed by the learners of GROUP "B" as they used more idioms, supplementary language chunks and other common phrases that the latter collection did not exploit. Additionally, their standpoints were not only more straightforward, these perspectives made more logical sense for the presented situations as well.

– What does "**discourse management**" mean for these social scientists? It means "the coherence, extent and relevance of each [learner's] individual contribution" to the oral communication. (Downes, 2004, p.57)

4) **Discourse Management:** Unlike many of the members of GROUP "B", the participants who were part of GROUP "A" made use of various practical terms combined with several keywords – lexis – which were linguistically apposite for this oral exam. Similarly, their knowledge of language rules – syntax – was dissimilar in terms of adequacy and effectiveness as it appeared to be richer and deeper.

– How do these scholars visualize "interaction"? They view it as "initiating and responding without undue hesitation, the ability to use interactive strategies to maintain or repair communication, and sensitivity to the norms of turn-taking." (Downes, 2004, p.57)

5) **Interaction:** In this regard, the students of GROUP "A" were able to maintain natural conversations, with common errors that did not really impede the communication amongst the interlocutors involved. It is the belief of these researchers that this trait alone turned these participants into subjects who felt certainly as well as surely prepared for this novel linguistic endeavor. Even though this specific set of learners handled themselves quite well before, during and after this oral exam, it must be duly noted that they seemed to be more in control, quite resourceful at and capable of handling the last segment of this verbal examination, without much hesitation.

Apart from those sets of findings, there are still others which these researchers have reckoned upon after doing this study and they will be mentioned in the subsequent lines, under different sub-headers to make their identification as well as their evident intrinsic relationships with this oral exam proposal a more natural task for the reader and a much more apparent deduction.

Design of the oral exam proposal – Designing the format of an oral exam which could be standardized for the use of thousands of students and dozens of language instructors, such as the one undertaken by this proposal, is an affair that should only be assumed and carried out by professionals in the EFL teaching field with qualifications that are similar (i.e. "a particular skill and training in test design" – Fulcher & Davidson, 2007, p.28) or superior (i.e. "people who specialize in test design" – Fulcher & Davidson, 2007, p.28) to the ones these researchers hold. If this explicit condition is not met, then those tasks produced for the tests will lack certain detailed properties and probably be as challenging and as useful as any plain classroom task.

The format upon which the Preliminary English Test (PET) Speaking Test from Cambridge functions has four distinctive parts (i.e. 1) each candidate interacts with the interlocutor, 2) simulated situation, 3) extended turn and 4) general conversation – Downes, 2004, p.55), but only two of those components (i.e. Part 2 and Part 3; in the suggested format, "individual turn" and "interactive turn" respectively) were adapted and modified to be used in this oral exam proposal. It is due to this pair of personalized, tailored items that these researchers were able to measure the spoken proficiency echelon of these students correctly; in other words, the design of the exam worked perfectly well towards achieving the goals of this research.

This oral exam proposal has also confirmed that "assessment is performance-based" (Fulcher & Davidson, 2007, p.29) when the arrangement of an examination is proper and well-organized. In fact, thanks to the format of this verbal assessment, these researchers were able to assess the performances of their students with a test that "encourages interaction, communication, … and [provides] feedback from … teacher to learner." (Fulcher & Davidson, 2007, p.29) It additionally endorses a "particular feature of the classroom context" which is "*collaboration* between learners." (Fulcher & Davidson, 2007, p.29)

In addition, for the format of this oral exam proposal to be successful, these scholars had to introduce their pupils to the two parts of the test (i.e. "individual turn" and "interactive turn"), recognize the actual linguistic needs of these learners, design suitable classes for their learning process, and inform them appropriately on the assessment instrument that would be exerted. In fact, providing key notes on what elements of a rubric are most difficult, most important, and most influential to get a fine grade is an instructional issue which has already been suggested effective by Petkov and Petkova (2006), and Reitmeier, Svendsen, and Vrchota (2004) who endorsed the notion that after "involving students in the … use of rubrics" these learners will probably show "improvements in academic performance." (Reddy & Andrade, 2010, pp.444-445) However, learning to administer and exploit rubrics is not just a job for the teachers; it is also for the learners. These pupils had to be well trained on the proper use of rubrics given that "simply handing out a rubric cannot be expected to have an impact on student work: students must be taught to actively use a rubric for self- and peer assessments and revision in order to reap its benefits." (Reddy & Andrade, 2010, p.445)

Finally, those learners that did well in the "interactive turn" of this verbal assessment abided by the five steps that should be followed in order to solve a problem using the Critical Thinking process:

Step 1 – I DENTIFY the Problem and Set Priorities.

Step 2 – D ETERMINE Relevant Information and Deepen Understanding.

Step 3 – E NUMERATE Options and Anticipate Consequence.

Step 4 – A SSESS the Situation and Make a Preliminary Decision.

Step 5 – S CRUTINIZE the Process and Self-Correct as Needed. (Facione, 2015, p.27)

Communicative Competence – Communicative competence was well-measured with this oral exam proposal as the students were required "to have more than [just] grammatical competence in order to be able to communicate effectively in [this foreign] language" (Susanto, 2012, p.7) while taking part in this verbal assessment. What is more, for increasing communicative competence, these instructors became "learning facilitators", were "creative and innovative" to include "thinking processes" during their class sessions, and instituted "the most effective teaching methods" for this set of students. (Abbaspour, 2016, p.149)

Fluency and Accuracy – Fluency (i.e. "the capacity to use language in real time, to emphasize meanings, possibly drawing on more lexicalized systems" – Ellis, 2009 in Abbaspour, 2016, p.147) as well as accuracy (i.e. "the ability to avoid error in performance, possibly reflecting higher levels of control in the language as well as a conservative orientation, that is, avoidance of challenging structures that might provoke error" – Ellis, 2009 in Abbaspour, 2016, p.147) were properly tested with this oral exam format.

Critical Thinking – CT (i.e. Critical Thinking) was developed in the class sessions that some of these students had during this research because of the "problem-solving" activities (Facione, 1990, p.16) that the CT Expert carried on with his group while preparing them for this oral exam proposal. As certifiable proof of development of CT in this set of learners, some of the students from this sample used "judgment, evaluation and argumentation to sift through opinions and arrive at those that [were] most valid" (Lai, 2011, p.29) while working in the "interactive turn" of this oral exam proposal.

Interaction / Interactive Communication – The training for the second part of this oral exam proposal allowed for this easily verifiable fact to happen:

“"Communication drives essentially from interaction" (Rivers, 1987, p. xiii).” (Abbaspour, 2016, p.147)

The “Interactive Communication” criterion from the analytic rubric featured on this oral exam proposal was entirely fulfilled every time that a student from this sample was able “to use language to achieve meaningful communication” (Downes, 2004, p.57) and that ideal situation was favorably recurrent right through the hours when this spoken assessment took place, especially with the group of the Critical Thinking (CT) Expert.

At the moment of doing the “interactive turn” of this oral exam proposal, the genuine interaction capability of the students from the CT Expert group showed no evidence nor influence of “unsupervised or unguided peer-interaction” (Abbaspour, 2016, p.147) since it was not allowed during their training. Therefore, it can be stated that fluency was not emphasized at the expense of accuracy whilst these learners were coached.

Rubrics – The rubrics developed for this oral exam proposal were “criterion-referenced” instead of “norm-referenced” as this attribute made them “more compatible with cooperative and collaborative learning environments” (Allen, 2014, p.2). Not only were these researchers able to gather oral positive feedback derived from the opinions of these students, these social scientists also declared that their own judgments towards the use of these new rubrics for this oral exam proposal were optimistic as well as encouraging. What is more, these constructive criticisms from these didactic stakeholders rested upon two easily recognizable facts: 1) “clarity and appropriateness of language [as] a central concern” (Reddy & Andrade, 2010, p.435) – fundamental for “the validity of rubrics”, and 2) these EFL instructors could practically assure that due to the clearness and correctness of this rubric, they would almost certainly grade other students performances with the same scores, in a future and probable scenery (i.e. “**rater reliability**” as expounded in Mackey & Gass, 2005, p.128) where this oral exam proposal were instilled.

The analytic criteria, made up by five specific categories (i.e. “Grammar and Vocabulary”, “Fluency”, “Pronunciation”, and “Interactive Communication”), applied to grade the performance of these students on this oral exam proposal were both suitable and relatable to the ones used in the “CEFR scale for spoken assessment” (i.e. “range, accuracy, fluency, interaction and coherence (Council of Europe, 2001: 28-29).”) (Roca, Varela & Palacios, 2013, p.61)

The use of analytic rubrics during this educational project indeed supported thinking and learning. As all of the students from the Critical Thinking Expert group had systematically acknowledged the criteria with which their spoken performances were going to be graded, they were able to perform quite well in this oral exam proposal. Additionally, because the management of this grading tool was accepted by all the members of this sample, both teachers were somewhat able to blend the concepts of instruction and assessment together. According to Andrade (2000, p.5), once that situation has taken place, as it did in this study, "the use of rubrics has a powerful effect on your teaching and, in turn, on your students' learning" too.

Instructional Implications – This study might serve as evidence that Critical Thinking (CT) skills and abilities can be developed and practiced in a normal classroom during an ordinary semester timetable as maintained by many critical thinking researchers. (Lai, 2011, p.29) Additionally, "collaborative or cooperative learning" helped the CT Expert a lot while teaching CT skills to his group because "students' relationships with others" are crucial in developing CT skills. (Lai, 2011, p.34) Similarly, these learners acquired "the ability to respond constructively to others during group discussion, which implies interacting in pro-social ways by encouraging and respecting the contributions of others." (Bailin et al. (1999) in Lai, 2011, p.34) Finally, the fact that the class sessions of the CT Expert were based on a constructivist perspective (i.e. "more student-centered than teacher-centered" (Bonk & Smith, 1998; Paul, 1992)) (Lai, 2011, p.36) was also truly helpful for the set objectives of this study.

Relationships to other concepts associated with EFL (English as a Foreign Language) teaching and learning – Motivation factors played a decisive and indisputable role for the Critical Thinking (CT) Expert group (i.e. intrinsic motivation – they wanted to improve their speaking skill) as well as for the Experienced Oral Examiner for International Exams collection (i.e. extrinsic motivation – they needed to pass the course) in support of their total commitment or as the main reason for their vague dedication, correspondingly, towards this study and its express requirements. (Lai, 2011, p.20)

When finding solutions for the problems of the "interactive turn", many of these students resorted to ""good", purposeful thinking (Paul and Elder, 2006)" (i.e. "Good thinking requires the ability to generate intellectual products" – Lai, 2011, p.21), which is the immediate and unequivocal result of putting creativity and CT together, as sustained by Lai (2011, p.21).

Unfortunately, all of these students felt anxiety during this oral exam proposal application. These researchers have made the educated assumption that these learners were worried about “being "wrong, stupid, or incomprehensible" (Brown, 2001, p. 269)”, which in turn “completely [affected their] speaking performance” (Abbaspour, 2016, p.146) hence their actual language production could have been faulty at times.

5.3 Conclusions

The main conclusions that these researchers made are all based upon the two central questions as well as the three sub-research questions that this educational project meant and needed to answer from the beginning.

First Central Question – What is the impact in EFL students of Pre-Intermediate level at an Ecuadorian state university on having their speaking skill evaluated with a standardized oral exam? The information gathered shows that the impact, regarding this specific issue, in all of these students is definitely positive. Likewise, after using this standardized oral exam proposal, the scores of many of these undergraduates certainly increased; in other words, those grades improved positively. Additionally, the majority of these students felt confident enough to detail that they were very well prepared for the proposed exam due to their proper training in the use of its rubric and their constant practice of suitable activities in class.

Second Central Question – What are the perceptions of EFL students of Pre-Intermediate level at an Ecuadorian state university on having their speaking skill evaluated with a standardized oral exam? A significant perception of a great number of these undergraduates is that they felt more comfortable taking this standardized oral exam proposal albeit having experienced anxiety all throughout (i.e. before, during, after) the examination process. Besides, highlighting the constructive features of this verbal assessment suggestion, the exit survey notified these social scientists that the majority of these students were clearly familiarized with (i.e. they really knew about) the process of the exam and the rubric or scale of evaluation. These pupils also acknowledged to being prepared in class with activities that would surely conduct them to perform successfully in this spoken test.

First Sub-Research Question – What is the level of approval that using this format of oral exam will have on these learners? In order to determine this insight, these students were given the chance to confirm or refute if they had felt more at ease with this new format in preference to any other layout that they had been assessed or valued with. The final results proved that a lot of these learners were content with being evaluated by means of the proposed verbal assessment design hence their discernment towards this innovative arrangement is of the assenting kind.

Second Sub-Research Question – What are the main advantages and disadvantages of using this format of oral exam that these learners perceived? For using this standardized oral exam proposal, these are the most important advantages distinguished by these pupils: 1) these learners were previously well-informed about the procedure and the rubric to be used to grade this verbal evaluation, 2) these students were thoroughly prepared in class (i.e. through strategies and activities) for the speaking test, and 3) even though teachers swapped groups in their exit oral exam, these undergraduates felt that the level of motivation from their evaluator was nearly the same as in their entry exam.

The main disadvantage that these learners recognized after utilizing this standardized verbal examination suggestion has to do with the anxiety they felt due to this oral exam; nevertheless, their nervousness was not incited by the proposed format. As a matter of fact, the level of anxiety before, while, and after the exam never changed in these undergraduates for this standardized oral exam proposal. What is more, these pupils still felt anxious notwithstanding knowing the procedure of the exam, being acquainted with the scale of evaluation (rubric), and having been trained with fitting activities. Therefore, it can be concluded that students always feel anxiety in an evaluation and that happens because they are genuinely worried about their real, definite performance.

Third Sub-Research Question – What specific recommendations will these learners make on this format of oral exam? These undergraduates felt that the number of activities developed in class to prepare them for this oral exam proposal was not insufficient; nonetheless, they will always need more practice. Hence, their first suggestion was providing students with adequate amounts of additional rehearsal time.

Additionally, these pupils recommended having more work in class with meaning focus speaking activities so as to improve their actual level of speaking proficiency.

There is another conclusion that has to be shared with the reader and it has to do with the washback effect that took place during this research. This stimulus was of the negative kind, and in addition to acutely modifying the original plan of action exclusively crafted for this project, it seriously shaped the final decisions made for this study. However, this negative washback was not provoked by these language instructors. In point of fact, it was the students under the training of the Experienced Oral Examiner for International Exams whom indicated this teacher their needs of finishing all the contents necessary for their final exam (i.e. a summative assessment composed by the topics and themes available on five different units) since they considered themselves not to be completely ready for their next course. After serious thought, this EFL language instructor caved in to the requests of those learners as he was also aware of the recurrent and severe linguistic deficiencies present in this specific set of students. Hence, this teacher could not follow the plan (i.e. set of tasks and activities) that had been previously designed by the Critical Thinking Expert, at all.

5.4 Recommendations

Working under the assumption that the academic authorities of this language center agree on using this innovative verbal assessment on their scholastic premises, here is an inventory of recommendations that will not just allow students to get better marks on this new oral exam, but teachers to provide their learners with better training for this distinctive spoken test.

Providing students with several opportunities to practice their exam management conduct (i.e. "what the candidate should say, and how they should behave, on entering and leaving the exam room" – Burgess & Head, 2005, p.137) as well as their speaking performance through mock tests will help learners become much more accustomed to this new format, in a shorter time, thereby allaying their "anxieties" since they will know "what the format of the exam will be, how the examiner will speak to them, and how they should respond to questions and instructions." (Burgess & Head, 2005, p.137) Furthermore, the researchers trust – based on empirical evidence provided by Pedley (2017, pp.10-11) regarding the causes of Foreign Language Anxiety (FLA) – that as

learners become habituated to being evaluated with the same format, their level of anxiety will decrease somehow.

Ashton and Thomas (2008, p.168) have allocated the language functions that will help candidates perform well on each part of PET speaking test. The first task of this proposal will certainly benefit the learners whose appliance of the functions used in **Speaking Part 2 (Discussing a situation)** (i.e. "Asking for and making suggestions", "Giving and explaining opinions", "Asking for/confirming opinions", "Agreeing", "Disagreeing", "Accepting that your opinion is different from someone else's") is suitable and satisfactory. The functions required for an apt and fitting performance in **Speaking Part 4 (Discussion)** (i.e. "Explaining/Asking for/Confirming Opinions" and "Asking about and expressing likes, dislikes and preferences") will definitely be useful to all of the students while they are doing the second task of this standardized oral exam proposal.

To balance accuracy and fluency while training for this oral exam proposal, the EFL instructors in charge of this tutoring should become aware that using the communicative approach during these mandatory coaching sessions will be the best choice as the learners' communicative competence will develop by encouraging them to "use relevant strategies in coping with certain language situations". (Abbaspour, 2016, p.148)

The EFL instructors who will be harnessing this novel format for assessing their students' speaking skill proficiency stage ought not to make the same mistake that the Experienced Oral Examiner for International Exams made, which was assuming that there will be no real, significant difference in the performances of his pupils and the ones from the Critical Thinking Expert – even if they had both received the same specific training that had to be carried out – due to the fact that these two groups were labeled as collections in need of ***"low-level exams"*** (Burgess & Head, 2005, p.121). In addition, his ill presumption that the complete lack of the planned tutoring aforementioned – which could actually come to pass in real life because of different events normal to this educational context, as it happened with his assembly – was but a minuscule and trivial drawback for the final results of this project had no supporting evidence that could be found anywhere in the qualitative data or the quantitative information provided by this study. As a matter of fact, the attained deductions and the verifiable figures of this final task point everything out towards the exact opposite direction. Therefore, no assumption regarding anything related to learners should ever

be made; instead, it would be best to establish a firm line of open and honest communication between the students and their teachers, so that they both share the same final, instructive aims.

Cambridge ESOL is one of many international bodies of assessment that keeps its representatives updated and trained by showing them "sample tests on video … selected to demonstrate … different levels of competence, [which] are pre-marked by a team of experienced assessors." (Downes, 2004, p.57) It would be ideal if this language center devoted some of its own resources to produce exactly the same. However, the present economic situation does not allow it to make such an investment hence these researchers would recommend doing series of intensive, practical training (i.e. proper application of rubrics) sittings a week or a weekend before the oral exams of this scholastic establishment take place with the intention of "calibrating" its teaching staff for this sturdy academic undertaking.

If this oral exam proposal is not accepted as a valid evaluation tool, then the next submission should attempt to measure both oral production and oral interaction, not just one of them. Additionally, the measurement of those speaking skill components should also be done separately, not together, just like it is done in tests that follow CEFR standards. (Roca, Varela & Palacios, 2013, p.58) Furthermore, that proposition must make use of "open-ended problem" (Lai, 2011, p.38) questions. It should also count with tasks that are not only "based on simulations that approximate real-world problems and issues" (Lai, 2011, p.39), but promote "exercise of judgment" (Lai, 2011, p.40) on the students too. Additionally, that suggestion should focus on the "quality of the arguments underlying [the] position" of the learners rather than on "the "correctness" [i.e. accuracy] of the answer." (Lai, 2011, p.40) Finally, apart from having an arrangement which serves the purpose of "reliability, or consistency" (Fulcher & Davidson, 2007, p.30) (i.e. inter-rater reliability), that offer should act in accordance with the "four assumptions" demanded on "large-scale testing" and denoted by Smith: "stability", "discrimination", "test length" and "homogeneity". (Fulcher & Davidson, 2007, p.31)

As another viable option, instead of using the oral exam proposal (i.e. its format, its tasks and its rubric) that these researchers crafted for this educational project, this Language Center could just try to get hold of previous, proven, established speaking tests from international institutions or prominent printing houses, along with their "assessment criteria" that will definitely be based upon "a number of standardised

scores and assessment statements which are ultimately linked to the six levels set out in the CEFR." (Roca, Varela & Palacios, 2013, pp.65-66)

5.5 Plan of action

These researchers future plan of action includes two specific matters: 1) to do a second study regarding this didactic concern (i.e. producing a much more effective and efficient oral exam format for the Language Center of this Ecuadorian state university), wielding the same procedure followed for this first research (i.e. employing the tasks originally prepared for Critical Thinking development on the sample, providing identical practical training on rubrics management for the EFL instructors involved in the project, applying this unaltered oral exam format as the exit trial, and using the exact same rubrics to grade students performances), except this time these scholars will not allow the sample to get divided as it unfortunately did in this initial attempt.

2) The researchers are aiming to generate a "test" that will be based on the initiatives conveyed by Gardner (1992) regarding a new approach to assessment. These pioneering suggestions embrace different notions and unique principles, such as

> Emphasis on assessment rather than testing.
>
> Assessment as simple, natural, and occurring on a reliable schedule.
>
> Ecological validity. (i.e. When individuals are assessed in situations which more closely resemble "actual working conditions," it is possible to make much better predictions about their ultimate performance.)
>
> Instruments which are "intelligence-fair". (i.e. The solution – easier to describe than to realize – is to devise instruments which are "intelligence-fair," which peer directly at the intelligence-in-operation rather than proceed via the detour of language and logical faculties.)
>
> Uses of multiple measures. (i.e. Attention to a range of measures designed specifically to tap different facets of the capacity in question … .)
>
> Sensitivity to individual differences, developmental levels, and forms of expertise.
>
> Use of intrinsically interesting and motivating materials.
>
> Application of assessment for the student's benefit. (pp.89-93)

In simple words, that examination will be the result of understanding, operating and handling an approach to evaluations that is completely different to the standardized testing fundamentals available on this proposal.

REFERENCES

Abbaspour, F. (2016). Speaking Competence and Its Components: A Review of Literature. *International Journal of Research in Linguistics, Language Teaching and Testing, 1*(4), 144-152.

Allen, M. J. (2014). *Using Rubrics to Grade, Assess, and Improve Student Learning.* Retrieved from October 23rd, 2017 from http://www.mdc.edu/sailearn/documents/4.1%20Rubric%20Workshop%20Handout-Mary%20Allen.pdf

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational leadership, 57*(5), 13-19.

Andrews, L. (2000). *Linguistics for L2 teachers.* Routledge.

Ashton, S., & Thomas, B. (2006). *PET Practice Tests Plus 2.* Pearson Education.

Blumenthal, E. (2010). *Introduction to SPSS 16.0* [PDF Document]. Retrieved March 11th, 2018 from http://julius.csscr.washington.edu/pdf/spss.pdf

Brown, H. D. (2004). *Language assessment. Principles and Classroom Practices.* White Plains, NY: Longman.

Burgess, S.& Head, K. (2005). *How to teach for exams.* Pearson Education Limited.

Carroll, R. T. (2007, January). Teaching critical thinking. In *Critical Thinking Workshop, Las Vegas.* Retrieved February 9th, 2018 from http://skepdic.com/essays/teachingCT.pdf

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education.* New York, NY: Routledge.

Council of Ministers of Education, Canada (CMEC) & Human Resources and Skills Development Canada (HRDSC). (2013). *Speaking for Excellence: Language Competencies for Effective Teaching Practice* [PDF Document]. Retrieved February 12th, 2018 from https://www.cmec.ca/Publications/Lists/Publications/Attachments/320/Speaking_for_Excellence.pdf

Creswell, J. W. (2015). *Educational research. Planning, conducting, and evaluating quantitative and qualitative research.* Pearson Education, Inc.

Deviant, S. (2017). *Cronbach's Alpha: Simple Definition, Use and Interpretation.* Retrieved March 11th, 2018 from http://www.statisticshowto.com/cronbachs-alpha-spss/

Downes, S. (2004). *PET Handbook, Sample Papers (Paper 3, Speaking)* [PDF Document]. Retrieved February 3rd, 2016 from https://www.cambridge-bayern.de/fileadmin/_migrated/content_uploads/PET_HB_sampleS_07.pdf

Dudovskiy, J. (2016). *Constructivism research philosophy.* Retrieved from http://research-methodology.net/research-philosophy/epistomology/constructivism/

educarplus.com. (2017). *Lista de Universidades del Ecuador por Categorías – 2018.* Retrieved March 5th, 2018 from http://educarplus.com/2017/03/lista-de-universidades-del-ecuador-por.html

Educational Testing Service. (2012, September). *Workshop Manual.* Propell Workshop for TOEFL iBT, Guayaquil, Ecuador.

Efstathiou, C. E. (2012). *Student's t-test: Comparison of two means.* Retrieved February 19th, 2018 from http://195.134.76.37/applets/AppletTtest/Appl_Ttest2.html

Elkind, D. (2005). Response to objectivism and education. *The Educational Forum, 69,* 328-334.

ELT World Wiki. (2017). *Backwash effect.* Retrieved September 30th, 2017 from http://elt.wikia.com/wiki/Backwash_effect

English Proficiency Test Online & Express Publishing. (2018). *English language levels (CEFR).* Retrieved February 25th, 2018 from https://tracktest.eu/english-levels-cefr/

Ennis, R. H. (1985). A logical basis for measuring critical thinking skills. *Educational leadership, 43*(2), 44-48.

Ennis, R. H. (2011). The nature of critical thinking: An outline of critical thinking dispositions and abilities. In *Sixth International Conference on Thinking, Cambridge, MA.* Retrieved February 12th, 2018 from http://faculty.education.illinois.edu/rhennis/documents/TheNatureofCriticalThinking_51711_000.pdf

Facione, P. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction. Research Findings and Recommendations.* Retrieved February 9th, 2018 from https://files.eric.ed.gov/fulltext/ED315423.pdf

Facione, P. A. (2015). *Critical Thinking: What It is and Why it Counts.* Retrieved February 9th, 2018 from https://www.insightassessment.com/About-

Us/Measured-Reasons/pdf-file/Critical-Thinking-What-It-Is-and-Why-It-Counts-PDF

Frederiksen, N. (1981). *The Real Test Bias.* (Report No. RR-81-40). Princeton, New Jersey: Educational Testing Service.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment.* London: Routledge.

Gardner, H. (1992). Assessment in context: The alternative to standardized testing. In *Changing assessments* (pp. 77-119). Springer, Dordrecht.

Gobierno Nacional de la República del Ecuador. (2013). *Objetivo 4. Fortalecer las capacidades y potencialidades de la ciudadanía. [4.8.i]* Retrieved from http://www.buenvivir.gob.ec/objetivo-4.-fortalecer-las-capacidades-y-potencialidades-de-la-ciudadania#tabs2

Grix, J. (2002). Introducing students to the generic terminology of social research. *Politics*, *22*(3), 175-186.

Harmer, J. (2007). *How to teach English.* Pearson Education Limited.

Herman, J. L., & Golan, S. (1993). The Effects of Standardized Testing on Teachers and Schools. *Educational Measurement: Issues and Practice*, *12*(4), 20-25.

İşisağ, K.U. & Demirel, Ö (2010). The Use of the Common European Framework of Reference for Languages on Developing the Speaking Skills. *Education and Science*, *35* (156), 190-204. Retrieved October 3rd, 2017 from https://search.proquest.com/docview/1009841837?pq-origsite=gscholar

Kelleghan, T., Madaus, G. F., & Airasian, P. W. (2012). *The effects of standardized testing* (Vol. 1). Springer Science & Business Media.

Lai, E. R. (2011) *Critical Thinking: A Literature Review. Research Report.* Retrieved February 12th, 2018 from https://images.pearsonassessments.com/images/tmrs/CriticalThinkingReviewFINAL.pdf

León, K. & Maldonado, J. (2017). *An Analysis of Teachers' Perceptions of Speaking Fluency Activities: Planning and Classroom Application.* Retrieved February 11th, 2018 from http://www.dspace.espol.edu.ec/xmlui/bitstream/handle/123456789/38402/D-CD216.pdf?sequence=-1&isAllowed=y

Lund Research Ltd. (2012). *Experimental mortality and internal validity.* Retrieved July 31st, 2017 from http://dissertation.laerd.com/internal-validity-p5.php

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Mianto, E. (2012). *Using rubrics to test students' performance speaking.* Retrieved October 3rd, 2017 from http://www.academia.edu/2205164/USING_RUBRICS_TO_TEST_STUDENTS_PERFORMANCE_IN_SPEAKING

Mostafa, N. A., & Othman, N. (2007). School-based Oral English Test-The Backwash Effect.

Nation, I. S., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking.* Routledge.

National Council of Teachers of English. (2010). *Standards for the Assessment of Reading and Writing: Standard 2.* Retrieved February 21st, 2018 from http://www.ncte.org/standards/assessmentstandards/standard2

NIST/SEMATECH. (2013). *e-Handbook of Statistical Methods.* Retrieved from https://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm

Paker, T. (2013). The backwash effect of the test items in the achievement exams in preparatory classes. *Procedia – Social and Behavioral Sciences, 70,* 1463-1471.

Paker, T., & Höl, D. (2012). Attitudes and perceptions of the students and instructors towards testing speaking communicatively. *Pamukkale University Journal of Education, 32*(2), 13-24.

Pedley, E. J. (2017). *The effect of gender and educational background on Foreign Language Anxiety (FLA) amongst EFL students at an Ecuadorian university*. (Unpublished master's dissertation). Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador.

Phelps, R. P. (2008). The role and importance of standardized testing in the world of teaching and training. Retrieved February 11th, 2018 from http://nonpartisaneducation.org/Review/Essays/v4n3.pdf

Prodromou, L. (1995). The backwash effect: from testing to teaching.

Puchta, H. (2012). *Developing thinking skills in the young learners' classroom.* Retrieved August 1st, 2017 from http://www.bing.com/cr?IG=66A140CAC2EC4764A1CBD5A17AB112BD&CID=23090CBD93A865701B55074E925564E6&rd=1&h=4AuWpPjpzdH6HtGMYOBNFj66htuirdMQ_2dJt3rVBwE&v=1&r=http://herbertpuchta.com/wp-content/files_mf/1337014114YL_Thinking_booklet.pdf&p=DevEx.LB.1,5507.1

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & evaluation in higher education*, *35*(4), 435-448.

Reid, S. (1998). Teaching critical thinking. *CELE (Center for English Language Education) Journal*, *6*, 17-22.

Roa, J. (2014). *Types of Language Assessment* [PowerPoint Slides]. Retreived September 10th, 2017 from https://es.slideshare.net/jhunarar/english-17

Roca-Varela, M. L., & Palacios, I. M. (2013). How are spoken skills assessed in proficiency tests of general English as a Foreign Language? A preliminary survey. *International Journal of English Studies*, *13*(2), 53-68.

Sayin, B. A. (2015). Exploring anxiety in speaking exams and how it affects students' performance. *International Journal of Education and Social Science*, *2*(12), 112-118.

Spratt, M., Pulverness, A., & Williams, M. (2005). *The TKT Teaching Knowledge Test Course.* Cambridge: CUP.

Susanto, J. (2012). Communicative Competence In Teaching Speaking. *Journal of English and Arabic Language Teaching*, *3*(1), 69-86.

Trinity College London. *Exam Information: Graded Examinations in Spoken English (GESE)* [PDF Document]. Retrieved February 12th, 2016 from http://www.trinitycollege.com/site/?id=368

Tsagari, D. (2007). *Review of Washback in Language Testing: What Has Been Done? What More Needs Doing?. Online Submission.* Retrieved February 10th, 2018 from https://files.eric.ed.gov/fulltext/ED497709.pdf

UCLA Institute for Digital Research and Education. (2017). *What does Cronbach's Alpha mean? | SPSS FAQ.* Retrieved March 11th, 2018 from https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/

University of Cambridge ESOL Examinations. (2011). *Using the CEFR: Principles of Good Practice.* Retrieved November 7th, 2016 fromhttp://www.cambridgeenglish.org/images/126011-using-cefr-principles-of-good-practice.pdf

University of Exeter. (2008). *Principles of assessment.* Retrieved September 10th, 2017 from http://www.exeter.ac.uk/staff/development/academic/resources/assessment/principles/types/

Wadham, R. (2011). *Proposal for the EBSS Instruction for Educators Committee to develop rubrics based on the ACRL Information Literacy Standards for Teacher Education for assessment purposes.* Retrieved October 23rd, 2017 from http://connect.ala.org/node/144089

Wall, D. (2000). The impact of high-stakes testing on teaching and learning: can this be predicted or controlled? *System*, *28*(4), 499-509.

APPENDIX 1 – Structured Questionnaire in Spanish

Cuestionario Estructurado para las Percepciones de los Estudiantes

Cuestionario realizado para la investigación de las percepciones de los estudiantes con respecto a la estandarización de los exámenes orales en una universidad estatal pública en Guayaquil-Ecuador.

Instrucciones: Por favor, señale con un visto (✓) y auto-evalúese honestamente basado en las percepciones que tuvo en todas las etapas de este examen oral - *antes*, *durante*, y *después*.

N	Preguntas	De Acuerdo	Parcialmente De Acuerdo	En Desacuerdo	Completamente En Desacuerdo
1	Estaba más ansioso y nervioso comparado a antes del examen oral.				
2	Estaba informado acerca del procedimiento del examen.				
3	Estaba muy cómodo antes del examen.				
4	No tenía idea alguna acerca de la escala de evaluación (rúbrica).				
5	Fui preparado en clases (a través de estrategias y actividades) para el examen oral.				
6	Tuve dificultad para expresarme claramente y en una forma adecuada durante el examen.				
7	Estaba ansioso y nervioso durante el examen oral.				
8	Fui alentado y motivado por el profesor durante el examen.				
9	Las tareas/actividades en el examen fueron de tipo que nunca antes había encontrado.				
10	Creo que las actividades orales en clases fueron insuficientes.				
11	Después del examen, me di cuenta que necesitaba más practica oral.				
12	Me sentí más cómodo haciendo este nuevo examen oral en vez de los anteriores.				

Muchas gracias por su cooperación.
Los Investigadores.

APPENDIX 2 – Structured Questionnaire in English

Structured Questionnaire for Students' Perceptions

Research made questionnaire on students' perceptions about oral standardized exams in a Public State University in Guayaquil-Ecuador.

Directions: Please check (✓) and rate yourself honestly based on the perceptions you had in all the stages of this oral exam - *pre*, *during*, and *post*.

N	Questions	Agree	Partially Agree	Disagree	Completely Disagree
1	I was more anxious and nervous compared to before doing the speaking test.				
2	I was informed about the procedure of the test.				
3	I was very comfortable before the test.				
4	I did not have any idea about the assessment scale (rubric).				
5	I was prepared in class (through strategies and activities) for the speaking test.				
6	I had difficulty in expressing myself clearly and in an adequate way during the test.				
7	I was anxious and nervous during the speaking test.				
8	I was encouraged and motivated by the teacher during the test.				
9	The tasks/activities in the test were types I had never encountered before.				
10	I believe that the speaking activities in classes were insufficient.				
11	After the test, I realized that I needed more speaking practice.				
12	I felt more comfortable doing this new oral exam rather than previous ones.				

Thank you very much for your cooperation.
The Researchers.

APPENDIX 3 – Current Oral Exam Format

CURRENT ORAL EXAM FORMAT

The existing oral exam is over 10 points and it considers these parameters: 1) *grammar and vocabulary*, 2) *fluency*, 3) *pronunciation* and 4) *interactive communication*. These four parameters have to be fully and exclusively aligned with the contents which have been revised throughout the course classes during the corresponding half of the term: first assessment *(mid-term exam)* and second assessment *(final term*).

The present oral exam is done in pairs or in groups of three, in exceptional cases. For the last scenario though, the maximum length of time for that oral exam is between seven and eight minutes. Although five minutes per pair is the maximum amount of time that a regular oral exam of this language center lasts, any oral exam could have a minimum duration of three minutes if the whole participation of the pair is *"excellent"* in all the aspects considered by present rubrics. Albeit the interaction pattern used for the currently offered oral exam, each student gets individual marks on the four parameters aforementioned hence undergraduates might do the oral exam tasks in pairs or groups of three, but their scores are based upon their personal performances with other interlocutor or interlocutors, not on the overall accomplishment of the pair or the group.

For the procedural component, all students are aware that they will have to produce a dialogue which comes from five different options as each one of the exams *(mid-term and final)* are comprised by contents available in the same number of previously selected units of the coursebook – five distinct alternatives per exam. During the semester, the teacher regularly informs learners on the particular tasks which will be selected for the oral exam, and on the class before its due date, the language instructor reminds the students about those topics once more. It is important to state that those topics are the basis for dialogues that have been frequently practiced in class beforehand, under the direct supervision of the teacher who must have provided formative, not summative, feedback during normal class sessions. No rehearsed dialogue is allowed therefore learners are required to generate and develop the formerly mentioned dialogues in class and in presence of their language instructor who uses the *Oral Exam Rubrics* designed by this language center in order to mark the contributions and participations of these students during their speaking assessments.

On oral exam day, each and every pair has the same preparation period (five minutes) as this amount of time is both adequate and applicable, considering the number of students that every group has – forty in total. Whilst a pair is doing the oral exam under

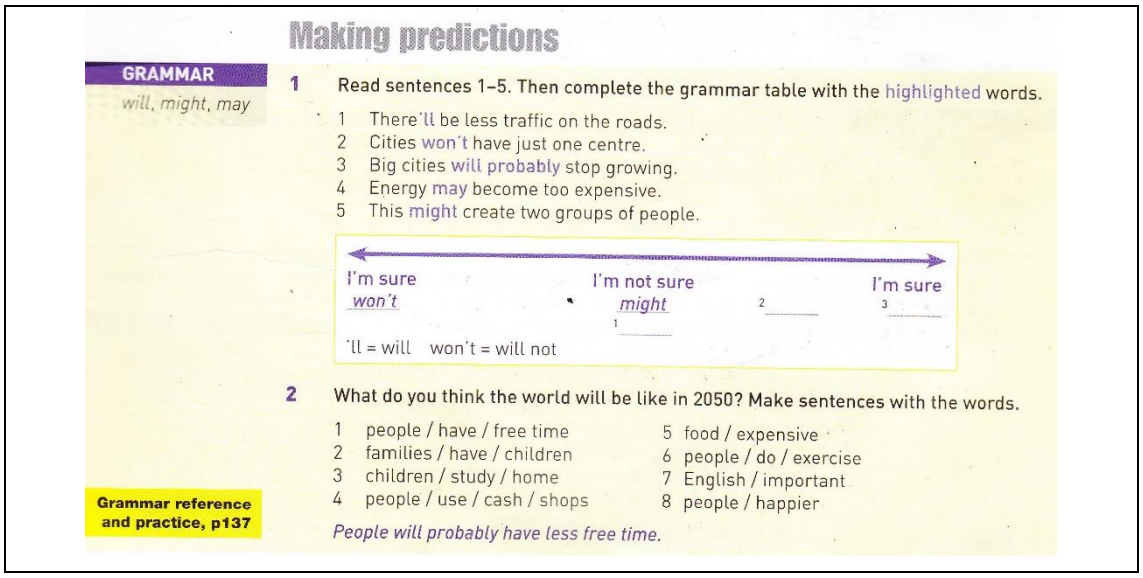
the scrutiny of the teacher, the next pair is preparing for their speaking assessment. At the end of a regular oral exam, the teacher gives each student a score which represents his or her performance for this summative evaluation. Once the speaking assessment is over, feedback provided to learners is minimum and immediate because students ought to have been exposed to proper use of rubrics during normal classes. As a final point, feedback is not given in any written form, it is only done orally.

CURRENT ORAL EXAM SAMPLE

Option	Language Function(s) related to Topic	Task Type	Task Format	Task Focus	Timing
1	Make guesses and predictions.	General conversation.	Learners interact with each other.	The learners talk together about their opinions, likes/dislikes, preferences, experiences, habits, etc.	3 to 5 minutes
2	Describe objects. Talk about possessions.	General conversation.	Learners interact with each other.	The learners give information of a factual, personal kind to their partners.	3 to 5 minutes
3	Say how you feel. Give and respond to different kinds of news. Ask for news.	General conversation.	Learners interact with each other.	The learners talk together about their opinions, likes/dislikes, preferences, experiences, habits, etc.	3 to 5 minutes
4	Talk about plans and arrangements. Make and change arrangements.	General conversation.	Learners interact with each other.	The learners use functional language to make and respond to suggestions, discuss alternatives, make recommendations and negotiate agreement.	3 to 5 minutes
5	Talk about homes and housing. Describe imaginary situations.	General conversation.	Learners interact with each other.	The learners talk together about their opinions, likes/dislikes, preferences, etc.	3 to 5 minutes

CURRENT ORAL EXAM ACTIVITIES SAMPLE

Option 1: Tilbury, A., Clementson, T., Hendra, L.A., Rea, D. & Doff, A. (2010). *ENGLISH UNLIMITED – B1 – Pre-Intermediate Coursebook:* Unit 7, page 59, exercise 2. Cambridge, University Press.



Making predictions

GRAMMAR
will, might, may

1 Read sentences 1–5. Then complete the grammar table with the highlighted words.

1 There'll be less traffic on the roads.
2 Cities won't have just one centre.
3 Big cities will probably stop growing.
4 Energy may become too expensive.
5 This might create two groups of people.

I'm sure	I'm not sure	I'm sure
won't	might	
	1 _____	2 _____
		3 _____

'll = will won't = will not

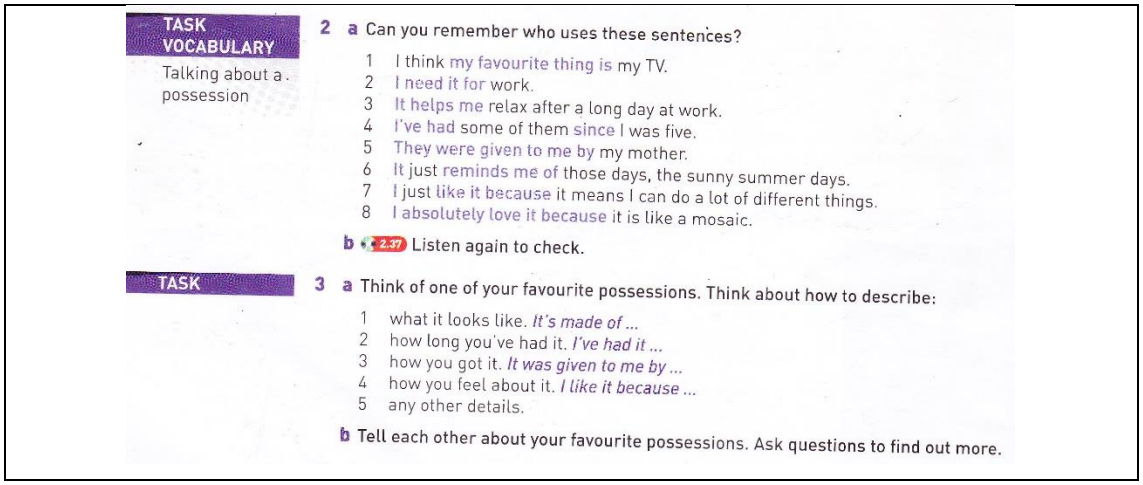
2 What do you think the world will be like in 2050? Make sentences with the words.

1 people / have / free time
2 families / have / children
3 children / study / home
4 people / use / cash / shops
5 food / expensive
6 people / do / exercise
7 English / important
8 people / happier

People will probably have less free time.

Grammar reference and practice, p137

Option 2: Tilbury, A., Clementson, T., Hendra, L.A., Rea, D. & Doff, A. (2010). *ENGLISH UNLIMITED – B1 – Pre-Intermediate Coursebook:* Unit 8, page 70, exercise 3. Cambridge, University Press.



TASK VOCABULARY
Talking about a possession

2 a Can you remember who uses these sentences?

1 I think my favourite thing is my TV.
2 I need it for work.
3 It helps me relax after a long day at work.
4 I've had some of them since I was five.
5 They were given to me by my mother.
6 It just reminds me of those days, the sunny summer days.
7 I just like it because it means I can do a lot of different things.
8 I absolutely love it because it is like a mosaic.

b 2.37 Listen again to check.

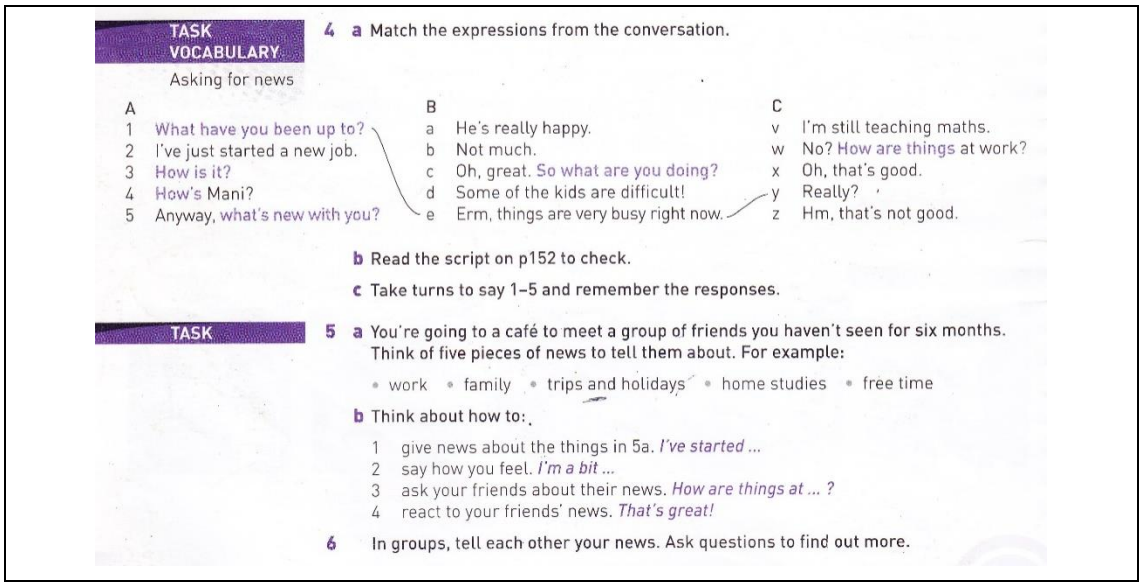
TASK

3 a Think of one of your favourite possessions. Think about how to describe:

1 what it looks like. *It's made of ...*
2 how long you've had it. *I've had it ...*
3 how you got it. *It was given to me by ...*
4 how you feel about it. *I like it because ...*
5 any other details.

b Tell each other about your favourite possessions. Ask questions to find out more.

Option 3: Tilbury, A., Clementson, T., Hendra, L.A., Rea, D. & Doff, A. (2010). *ENGLISH UNLIMITED – B1 – Pre-Intermediate Coursebook:* Unit 9, page 78, exercises 5 & 6. Cambridge, University Press.



TASK VOCABULARY
Asking for news

4 a Match the expressions from the conversation.

A		B		C	
1	What have you been up to?	a	He's really happy.	v	I'm still teaching maths.
2	I've just started a new job.	b	Not much.	w	No? How are things at work?
3	How is it?	c	Oh, great. So what are you doing?	x	Oh, that's good.
4	How's Mani?	d	Some of the kids are difficult!	y	Really?
5	Anyway, what's new with you?	e	Erm, things are very busy right now.	z	Hm, that's not good.

b Read the script on p152 to check.

c Take turns to say 1–5 and remember the responses.

TASK

5 a You're going to a café to meet a group of friends you haven't seen for six months. Think of five pieces of news to tell them about. For example:

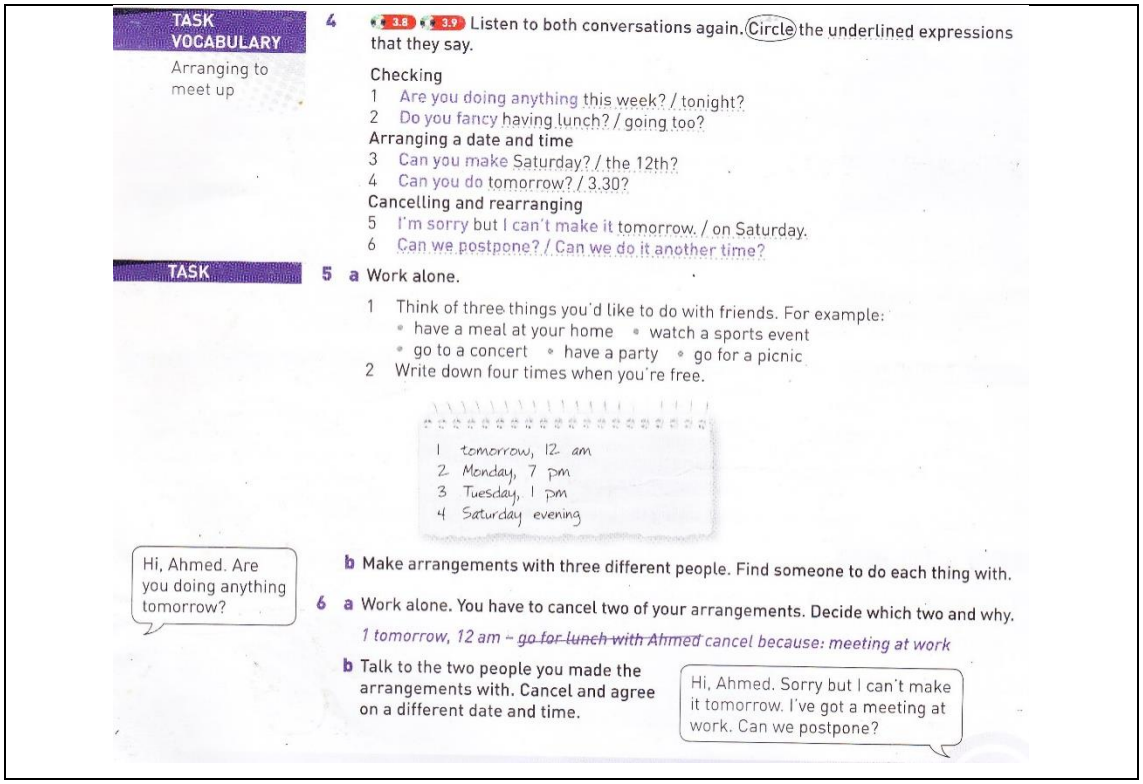
• work • family • trips and holidays • home studies • free time

b Think about how to:

1 give news about the things in 5a. *I've started ...*
2 say how you feel. *I'm a bit ...*
3 ask your friends about their news. *How are things at ... ?*
4 react to your friends' news. *That's great!*

6 In groups, tell each other your news. Ask questions to find out more.

Option 4: Tilbury, A., Clementson, T., Hendra, L.A., Rea, D. & Doff, A. (2010). *ENGLISH UNLIMITED – B1 – Pre-Intermediate Coursebook:* Unit 10, page 86, exercises 5 & 6. Cambridge, University Press.



TASK VOCABULARY
Arranging to meet up

4 3.8 3.9 Listen to both conversations again. Circle the underlined expressions that they say.

Checking
1 Are you doing anything this week? / tonight?
2 Do you fancy having lunch? / going too?
Arranging a date and time
3 Can you make Saturday? / the 12th?
4 Can you do tomorrow? / 3.30?
Cancelling and rearranging
5 I'm sorry but I can't make it tomorrow. / on Saturday.
6 Can we postpone? / Can we do it another time?

TASK

5 a Work alone.

1 Think of three things you'd like to do with friends. For example:
• have a meal at your home • watch a sports event
• go to a concert • have a party • go for a picnic
2 Write down four times when you're free.

1 tomorrow, 12 am
2 Monday, 7 pm
3 Tuesday, 1 pm
4 Saturday evening

Hi, Ahmed. Are you doing anything tomorrow?

b Make arrangements with three different people. Find someone to do each thing with.

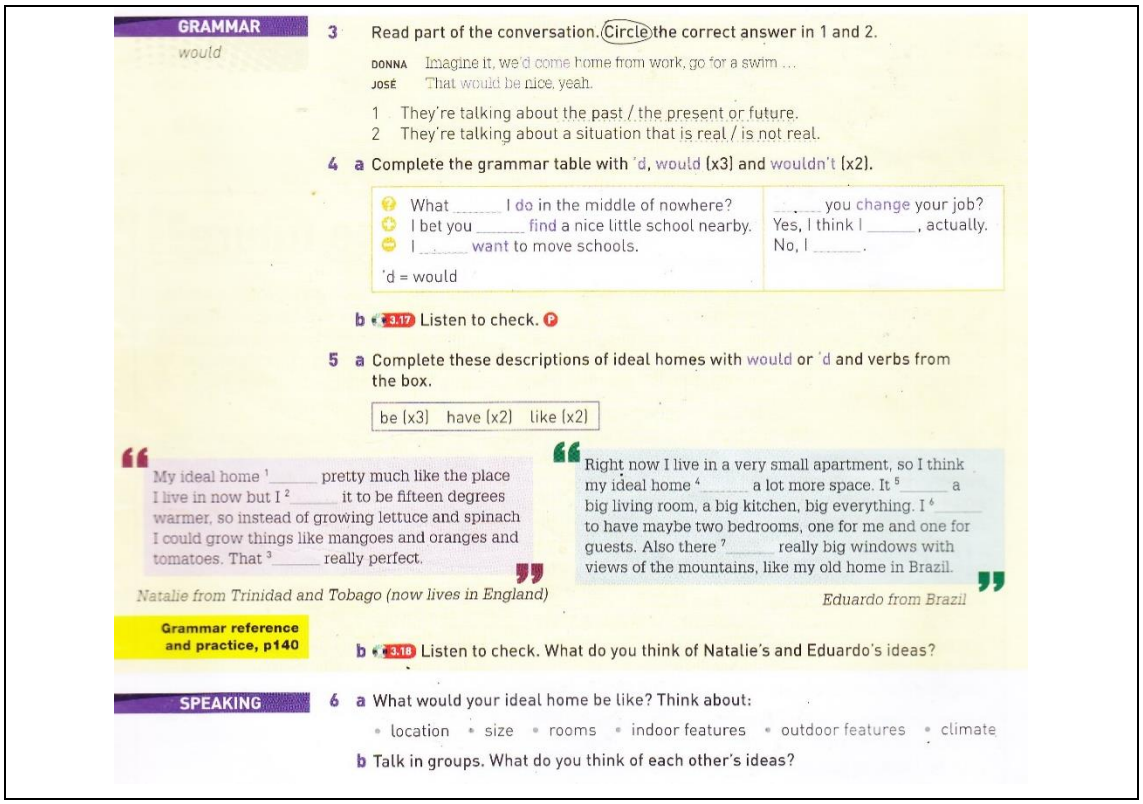
6 a Work alone. You have to cancel two of your arrangements. Decide which two and why.

1 tomorrow, 12 am – ~~go for lunch with Ahmed~~ cancel because: meeting at work

b Talk to the two people you made the arrangements with. Cancel and agree on a different date and time.

Hi, Ahmed. Sorry but I can't make it tomorrow. I've got a meeting at work. Can we postpone?

Option 5: Tilbury, A., Clementson, T., Hendra, L.A., Rea, D. & Doff, A. (2010). *ENGLISH UNLIMITED – B1 – Pre-Intermediate Coursebook:* Unit 11, page 91, exercise 6. Cambridge, University Press.



GRAMMAR *would*

3 Read part of the conversation. Circle the correct answer in 1 and 2.

DONNA Imagine it, we'd come home from work, go for a swim ...
JOSÉ That would be nice, yeah.

1 They're talking about the past / the present or future.
2 They're talking about a situation that is real / is not real.

4 a Complete the grammar table with 'd, would (x3) and wouldn't (x2).

What _____ I do in the middle of nowhere? I bet you _____ find a nice little school nearby. I _____ want to move schools. 'd = would	_____ you change your job? Yes, I think I _____, actually. No, I _____.

b 3.17 Listen to check. P

5 a Complete these descriptions of ideal homes with would or 'd and verbs from the box.

be (x3) have (x2) like (x2)

"My ideal home [1]_____ pretty much like the place I live in now but I [2]_____ it to be fifteen degrees warmer, so instead of growing lettuce and spinach I could grow things like mangoes and oranges and tomatoes. That [3]_____ really perfect."

Natalie from Trinidad and Tobago (now lives in England)

"Right now I live in a very small apartment, so I think my ideal home [4]_____ a lot more space. It [5]_____ a big living room, a big kitchen, big everything. I [6]_____ to have maybe two bedrooms, one for me and one for guests. Also there [7]_____ really big windows with views of the mountains, like my old home in Brazil."

Eduardo from Brazil

Grammar reference and practice, p140

b 3.18 Listen to check. What do you think of Natalie's and Eduardo's ideas?

SPEAKING

6 a What would your ideal home be like? Think about:

• location • size • rooms • indoor features • outdoor features • climate

b Talk in groups. What do you think of each other's ideas?

APPENDIX 4 – Analytic Rubric for Current Oral Exam Format

Intermediate B Oral Exam Rubric

1. Students must take the oral exam in pairs or groups.
2. The tasks must include at least one activity where the students interact with each other without intervention by the examiner.
3. The tasks must assess the following learning outcomes:
   - ❖ Students can enter unprepared into conversations on topics that are familiar, of personal interest or relevant to their everyday life. They can give simple reasons and explanations for their opinions and plans.
   - ❖ Students can narrate their own personal experiences, and describe reactions and feelings with some prior preparation. (production)

Note: the topics that are chosen for the mid-term and final oral exams respectively should be based on what has been taught in that half of the course and the overall learning outcomes.

	0	1	2	3
Grammar and Vocabulary	The student does not appropriately use structures or vocabulary items and makes many impeding errors.	The student appropriately uses structures and vocabulary with some non-impeding errors, but the language is significantly more basic than the level of the course. OR The student attempts to use a limited range of structures and vocabulary from the course, but makes some impeding errors.	The student appropriately uses a limited range of structures and vocabulary from the course with some non-impeding errors. OR The student attempts to use a wide range of structures and vocabulary from the course, but makes a few impeding errors.	The student appropriately uses a range of structures and vocabulary from the course with some non-impeding errors.

Fluency	The student cannot form more than one phrase or sentence without long periods of hesitation.	The student can form sentences together and react to most questions, but hesitation often impedes conversation.	The student can form several sentences together and react to most questions, but occasionally hesitates for an unnatural length of time.	The student can form several sentences together and react to questions without any undue hesitation.
Pronunciation	The student cannot be understood.	The student can sometimes be understood and sometimes not.	The student can be understood on the whole, but certain words are difficult to understand or the speech sounds very unnatural.	The student can be easily understood and shows fairly natural speech (intonation, stress and linking).
Interactive communication	The student cannot interact with his/her partner.	The student can respond, but not initiate interaction.	The student can respond and occasionally initiates interaction.	The student can respond and initiate interaction.
Global achievement	Very bad for their level.	Weak for their level.	Satisfactory for their level.	Good or excellent for their level.

APPENDIX 5 – Oral Exam Proposal

ORAL EXAM SCRIPT:

TEACHER (T): Good morning / Good afternoon. My name is *Teacher's name.*

Looking at A: What is your name?

STUDENT A (A): XXX / My name is…

* T writes down A's name on the piece of paper with the rubric to be used for grading A's performance.

T: OK, thank you.

Looking at B: And what is your name?

STUDENT B (B): XXX / My name is…

* T writes down B's name on the piece of paper with the rubric to be used for grading B's performance.

T: OK, thank you.

Looking at A and B: Before we start the oral exam, let me tell you that this test has **two parts**. In **part one**, you are going to **work individually** and in **part two** you are going to **work in pairs**. OK? Let's start.

T: Looking at A and B: **Part One:** In this part, each of you is going to **answer a question** and you have about **one minute** for that.

Looking at A: Student A's name. * T chooses and asks any *one* of the questions from the "Individual Turn" section available below.

A answers the question. T listens to A's response and pays attention to time. *No answer should last more than one minute.* While A answers, T is grading A's performance using the rubric.

Looking at A: Okay, thank you.

Looking at B: Student B's name. * T chooses and asks any *one* of the questions from the "Individual Turn" section available below.

B answers the question. T listens to B's response and pays attention to time. *No answer should last more than one minute.* While B answers, T is grading B's performance using the rubric.

Looking at B: Thank you.

Looking at A and B: Now, let's continue. **Part Two:** For this second part, you are going to **work together**. I am going to give you **a situation**. You are going to talk about it and then, **make a decision**. You have about **two minutes** for this part. OK? Let's start.

* T chooses and uses any *one* of the situations from the "Interactive Turn" section available below. T hands in the situation printed on the booklet and *reads the situation to the students aloud from his / her script*. T makes emphasis on the **bolded instructions** so A and B realize their importance.

A and B must use critical thinking to work out and around the situation. A and B have to interact among each other using appropriate syntax (grammar) and lexis (vocabulary). T listens to A and B's interaction and pays attention to time. *No answer should last more than two minutes.* While A and B are interacting, *T is grading each person's individual – not the pair's – performance* using the rubric.

Looking at A and B: Okay, thank you very much. That is the end of the test.

INDIVIDUAL TURN:

Unit 12

1) Tell me about a person or a member of your family that you admire and why.

2) Tell me about a place in Ecuador or in South America that you would like to visit and why.

3) Tell me about the city or town where you grew up. Give me a short description of what growing up there (in that city or town) was like.

Unit 13

1) Tell me about the electronic gadget that you use the most or that you use daily. Why do you use it so often? How do you feel about it?

2) Tell me about the activities that you used to do when you were a child / in school / in high school.

3) Tell me about the places that you used to go to on vacations when you were a child / in school / in high school.

4) Tell me about how using technology has changed your life. Describe how using a smartphone or a computer has made your student life easier.

5) "Ecuadorians are not always very punctual." Why do you think that is?

6) Tell me about your first cell phone or computer. How and when did you get it? What was it like? How often did you use it? How did you feel about it then? How do you feel about it now?

Unit 14

1) Do you think that: "Everyone should learn at least two languages."? Why? / Why not?

2) Do you think that: "People should only worry about their health if they drink and smoke."? Why? / Why not?

3) What things can we do to reduce our carbon footprint? Name at least three activities and explain them.

4) How can a person improve his / her English skills? Name at least three activities and explain them.

Unit 1

1) Tell me about your favorite TV show. Why do you like it so much?

2) Tell me about an important piece of news that has happened recently in our country / in our city.

3) To get your news, what do you prefer: the Internet, newspapers, or the TV? Why?

4) Do you think that: “ESPOL is the best place to study and get a degree in Guayaquil / in Ecuador.”? Why? / Why not?

5) Tell me about your favorite movie / the last movie you saw / any movie you have seen recently. Did you enjoy it? Why? / Why not?

6) Imagine a friend from a different country wants you to tell him about Ecuador and its customs – eating habits, family life, hospitality, etc. What will you tell him?

Unit 2

1) What social networks do you have: Twitter, Facebook, Instagram? What is your opinion about them? Why do you use them? How often do you use them?

2) Do you think that: “It is better to talk face to face with a person that you have a problem with instead of talking about it over the phone.”? Why? / Why not?

3) How much time do you spend online? What sites do you use most? Why?

4) Do you think that: “Children in the future are more likely to learn how to write a language in a computer than on a piece of paper.”? Why? / Why not?

5) Do you think that: “Printed newspapers and books will definitely disappear in the future.”? Why? / Why not?

6) Do you think that: “You will probably move to another country in the next ten years.”? Why? / Why not?

7) Do you think that: “The use of mobile phones in hospitals should be banned / prohibited if doctors use them while taking care of patients.”? Why? / Why not?

8) Do you think that: “Studying at ESPOL is so difficult because the teachers tend to send too much homework.?” Why? / Why not?

9) Do you think that: “Real relationships can start from online relationships.?” Why? / Why not?

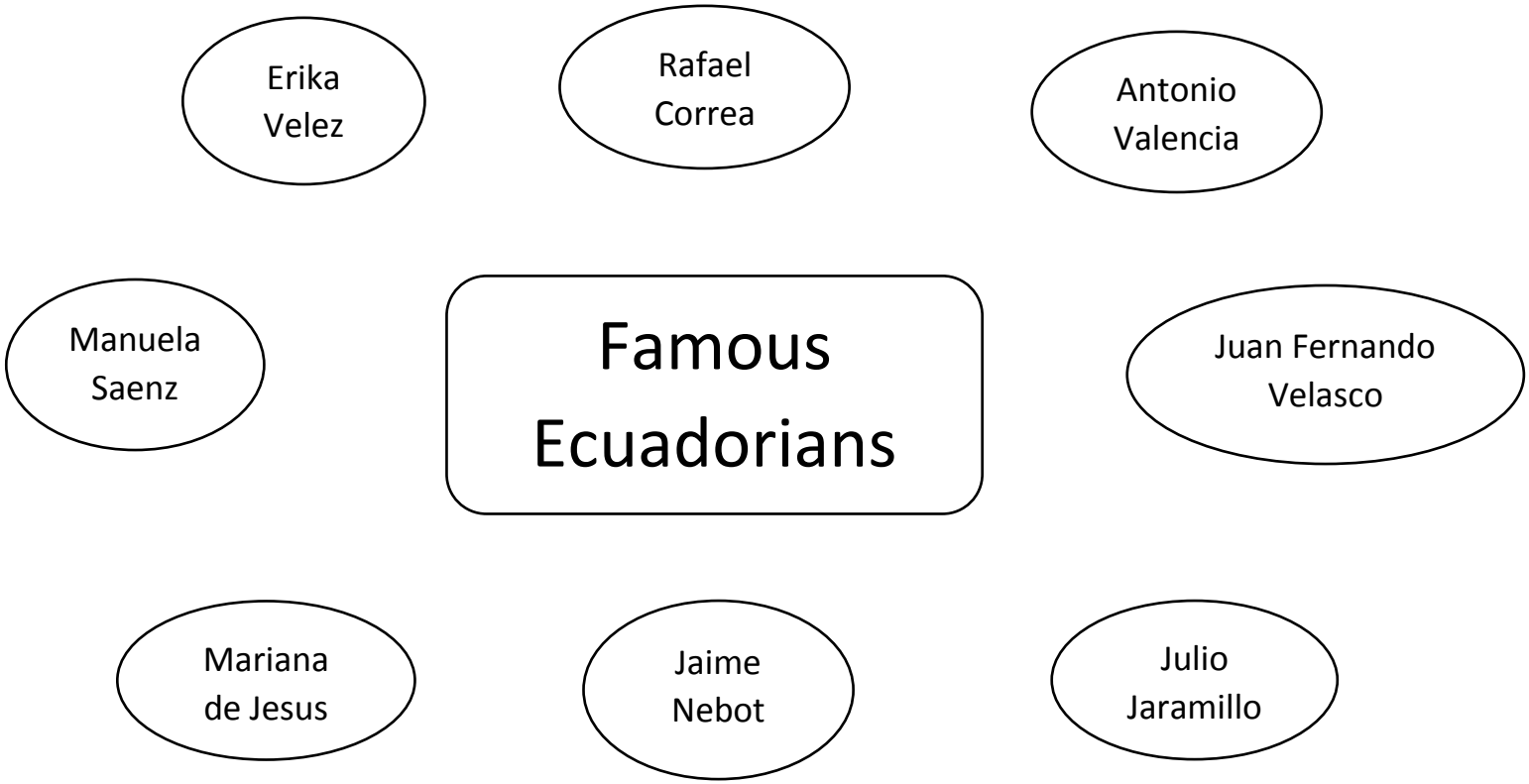
INTERACTIVE TURN:

Unit 12

You want to visit **two cities or towns** in Ecuador during this Carnival. Talk about **why** you think **those two cities or towns will be great to visit** and what **activities** you would like to do there during Carnival.

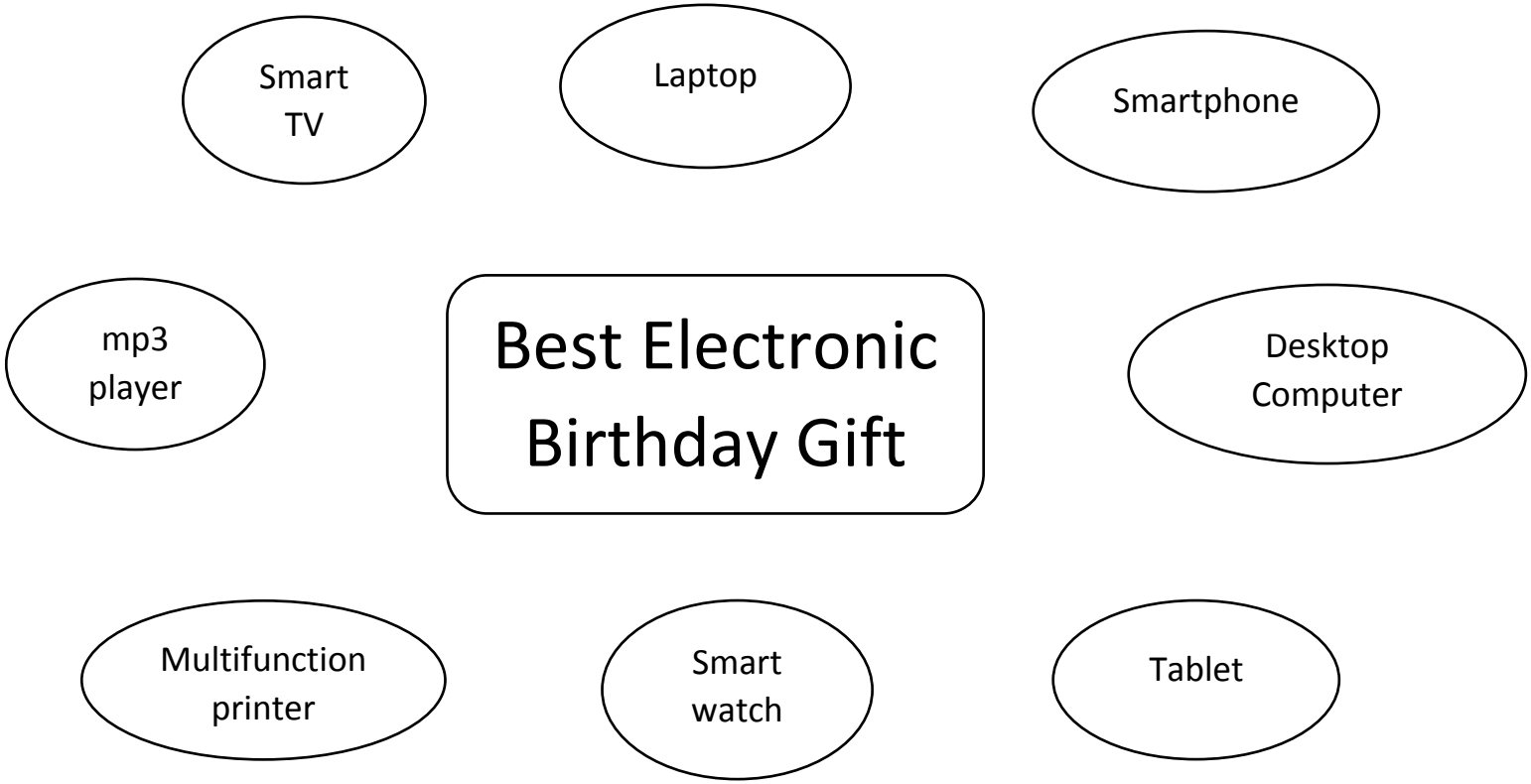
Unit 12

Your English teacher wants you to make an **oral presentation about a famous person from Ecuador**. Decide **which famous person will be interesting** for this presentation and the **type of information** you will include in the presentation.



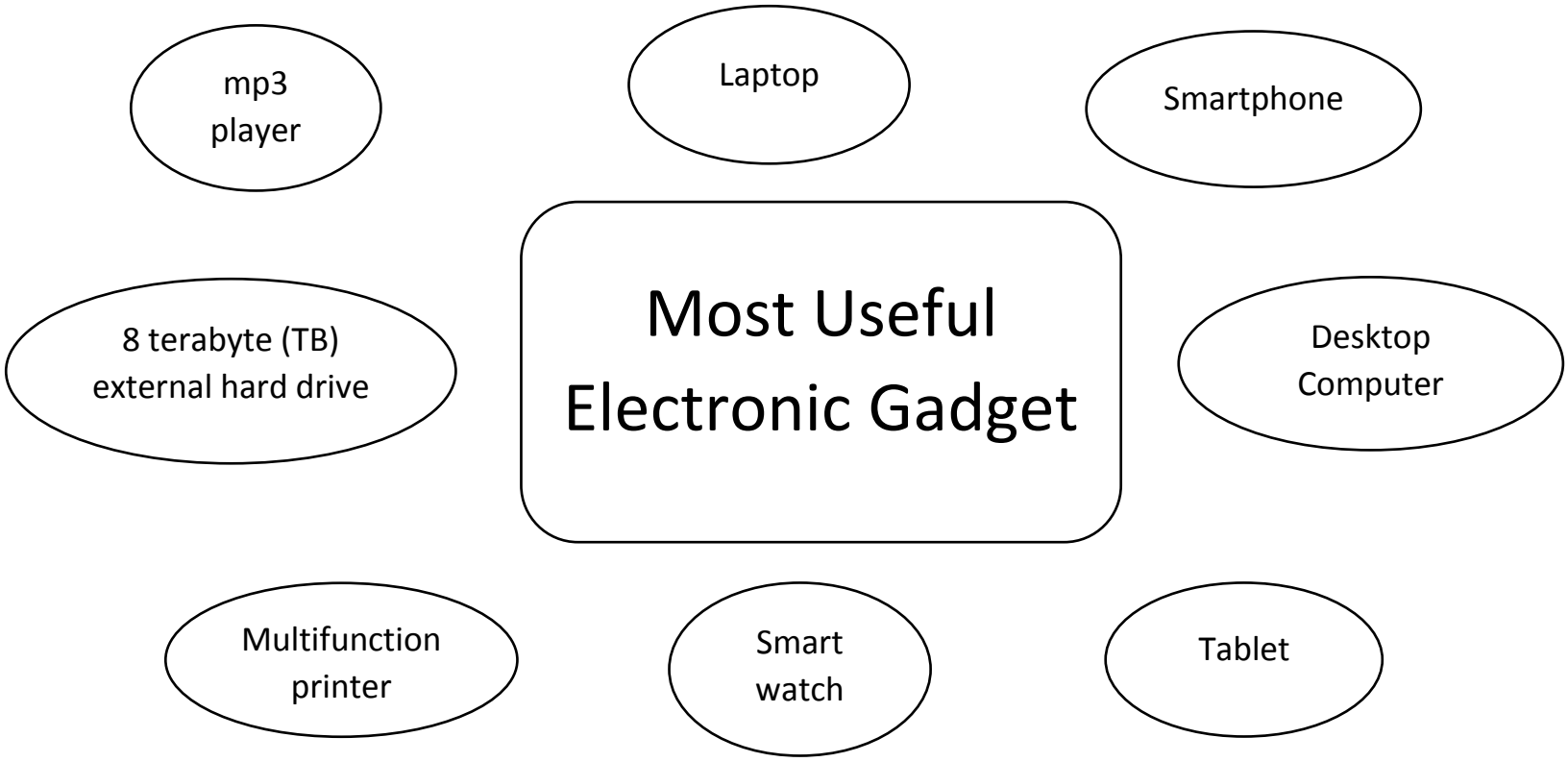
Unit 13

The **birthday of a really good friend** of yours is next week. You and your partner are going to decide **which of these electronic gadgets will be the best gift** for your good friend. Indicate **why** you chose **that electronic gadget over the others**.



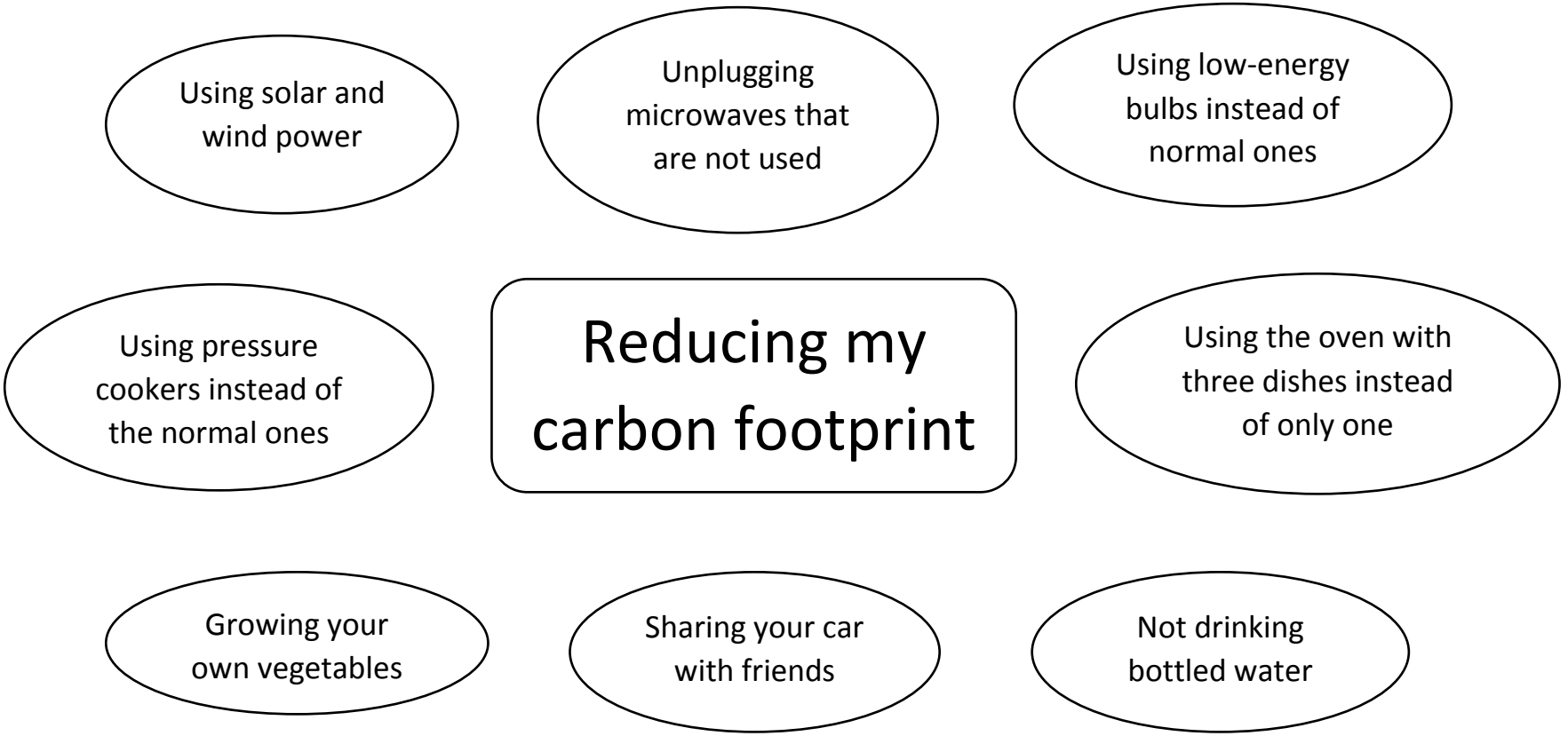
Unit 13

The Mayor of Guayaquil wants to give the best high school students of the city **an electronic gadget that will be the most useful** for them **during university**. Decide **which of these gadgets will be the best choice**. Indicate **why** you chose **that electronic gadget over the rest.**



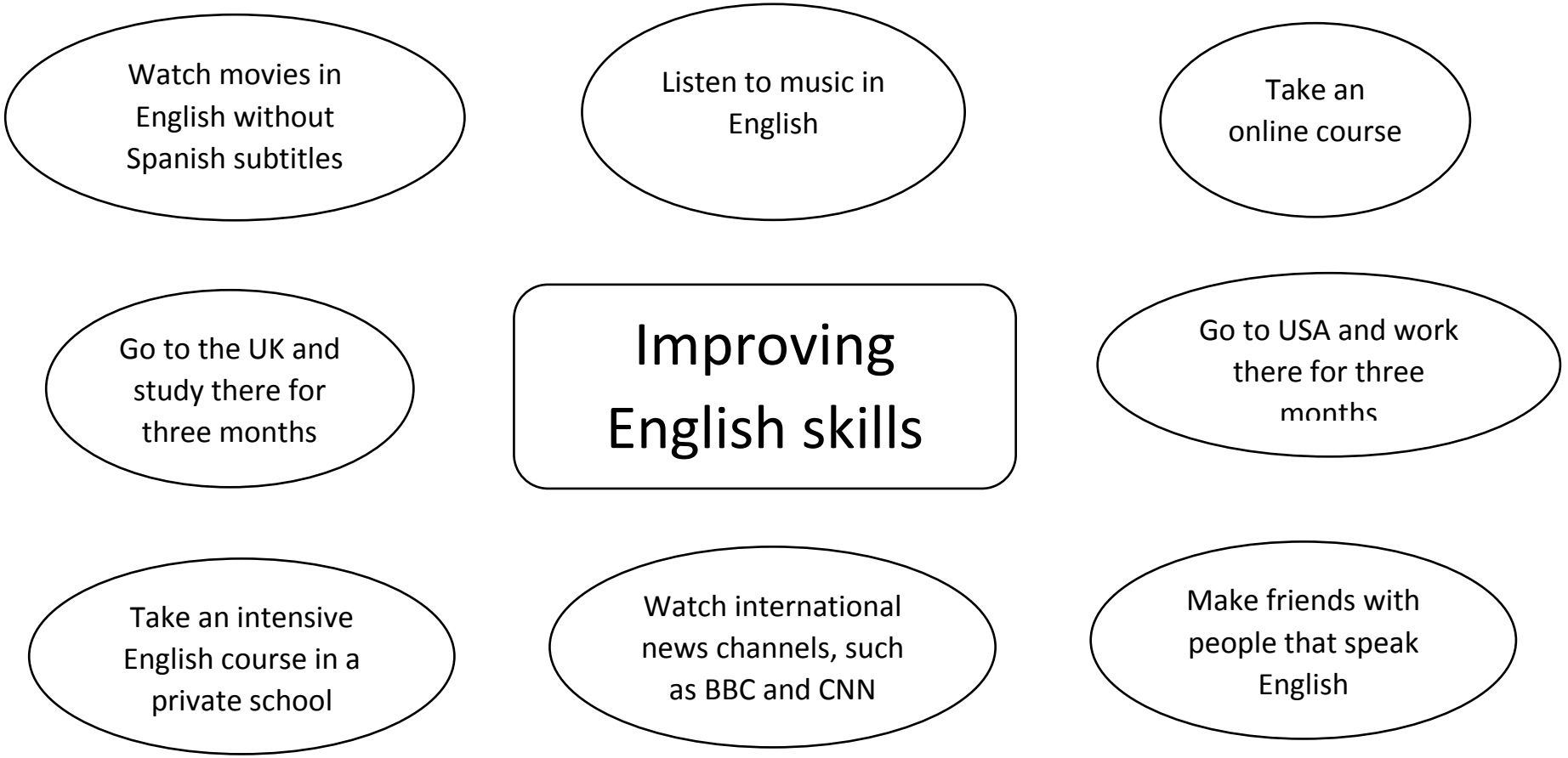
Unit 14

Here are some **common activities** that **people do to reduce their carbon footprint**. Decide **which three will be best to try** and **why**.



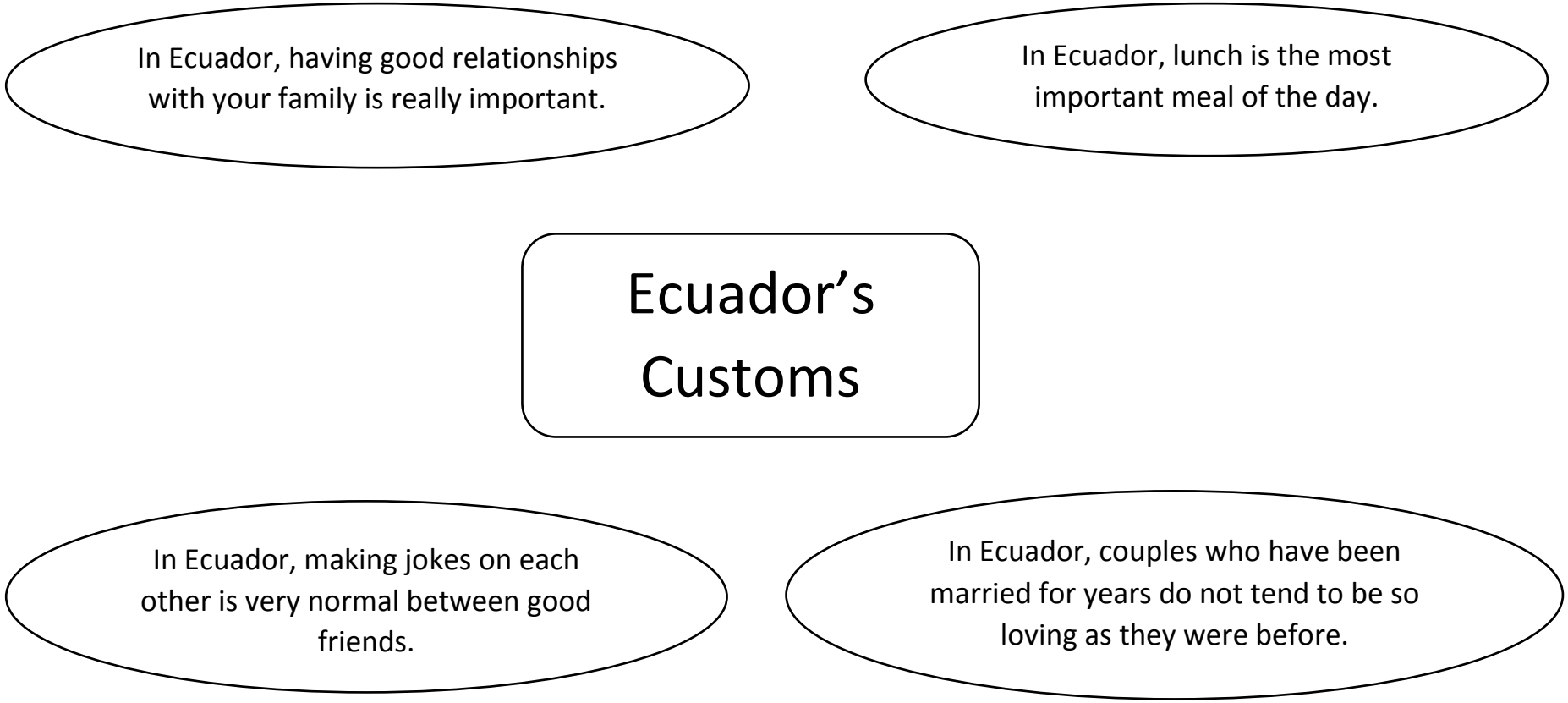
Unit 14

Here are some **common ways to improve a person's English skills**. Decide **which three will be best to try** and **why**.



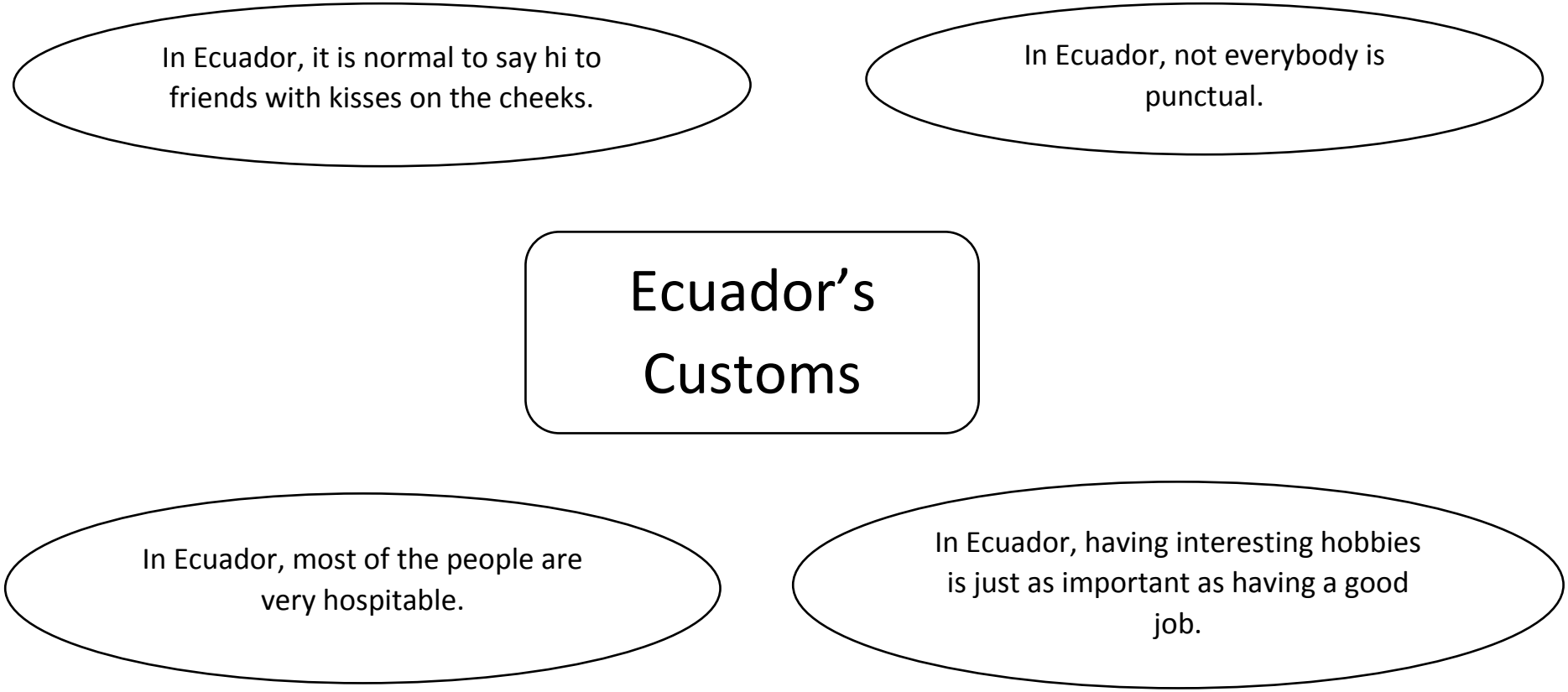
Unit 1

Your English teacher wants you to make a **short video**, which will be uploaded onto YOUTUBE, **about Ecuador and its customs**. Decide **which two** of these **aspects have to be included in the video** and **why**.



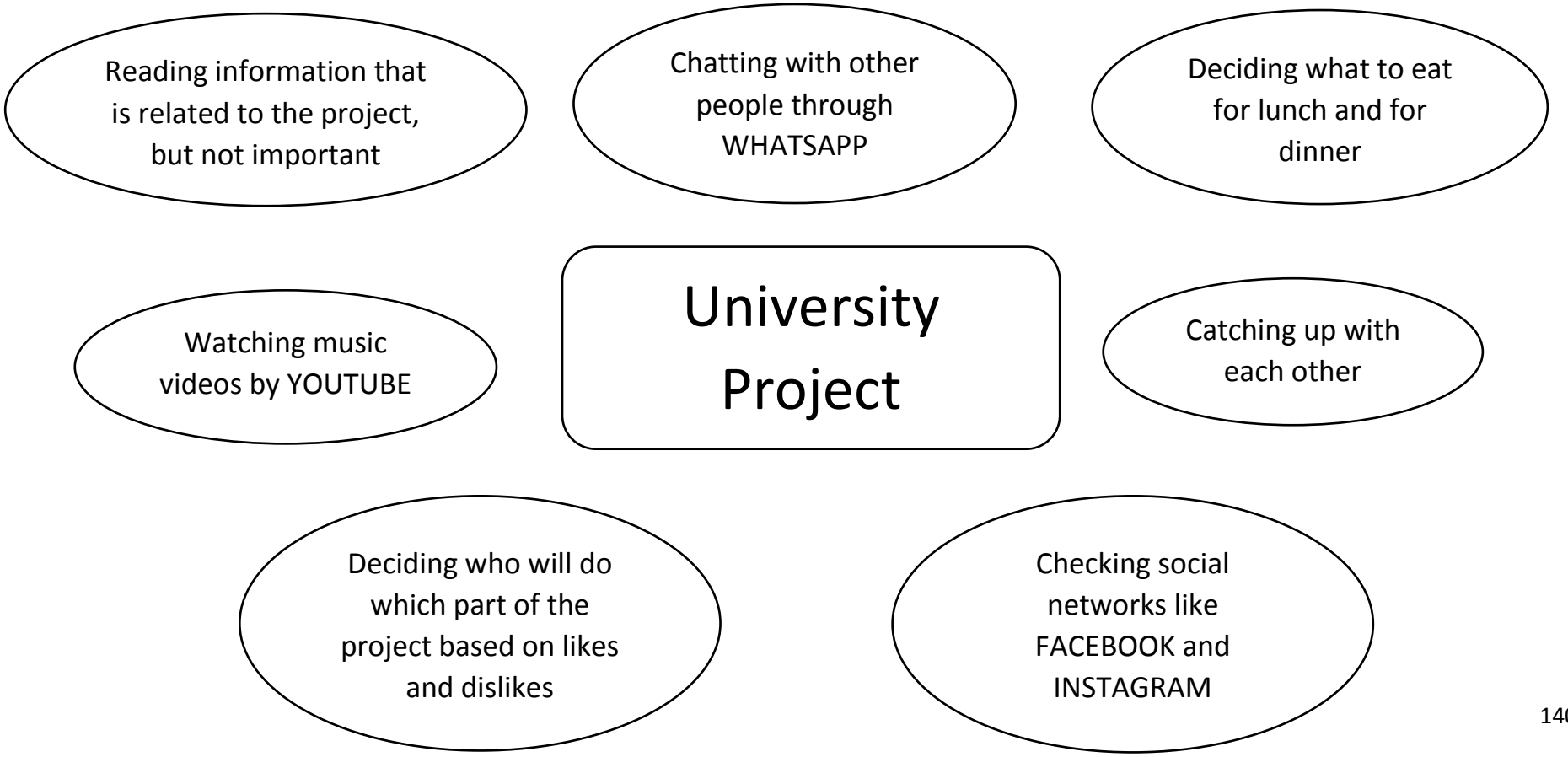
Unit 1

Your English teacher wants you to make a **short video**, which will be uploaded onto YOUTUBE, **about Ecuador and its customs**. Decide **which two** of these **aspects have to be included in the video** and **why**.



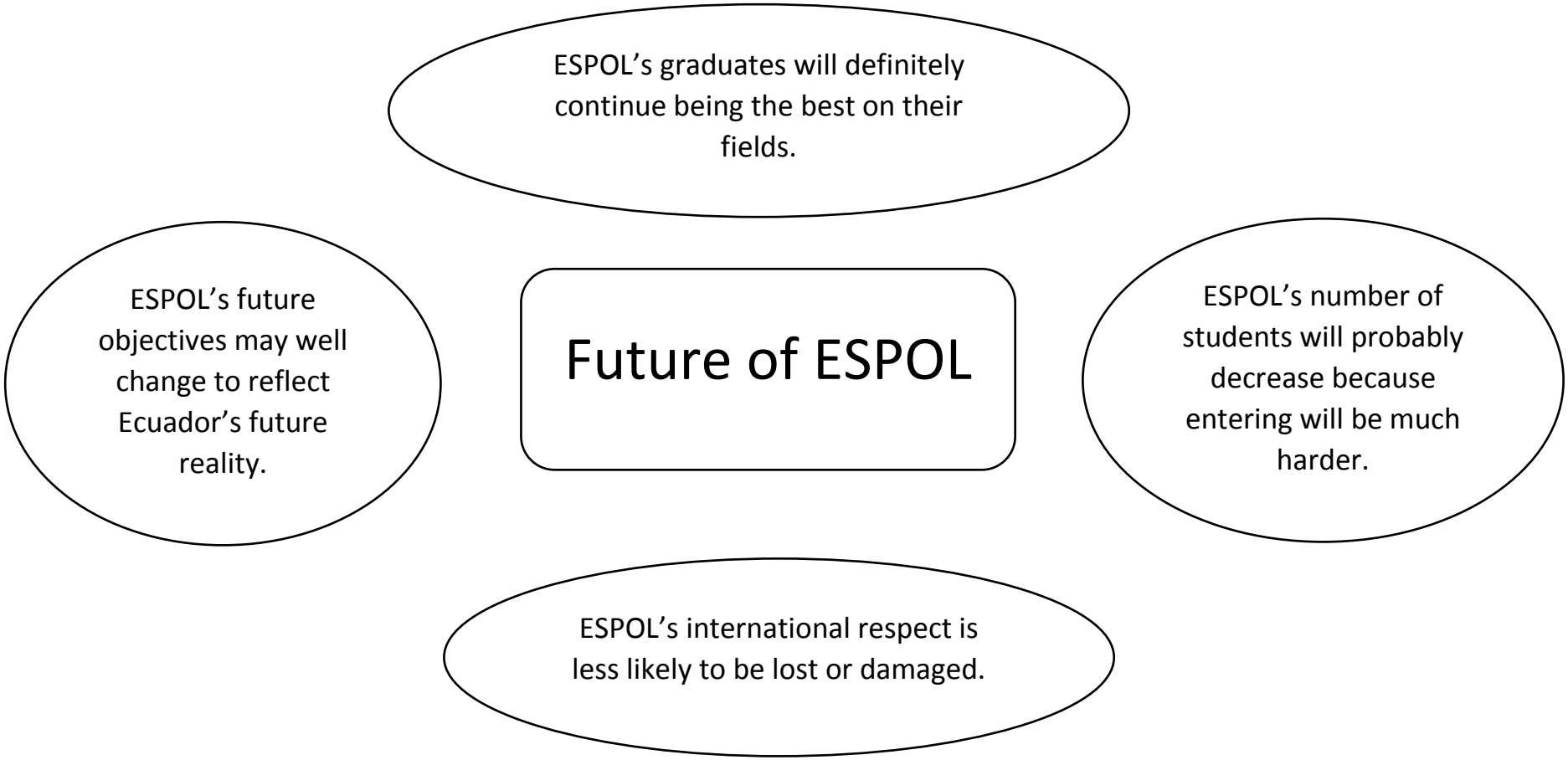
Unit 2

You have to do **a really important university project for tomorrow**. **Today**, you are **getting together for doing that right after this exam**. Decide **which of these usual distractions you will not stand** and **why** so that you can finish the project on time. Indicate **how you will avoid those distractions**.



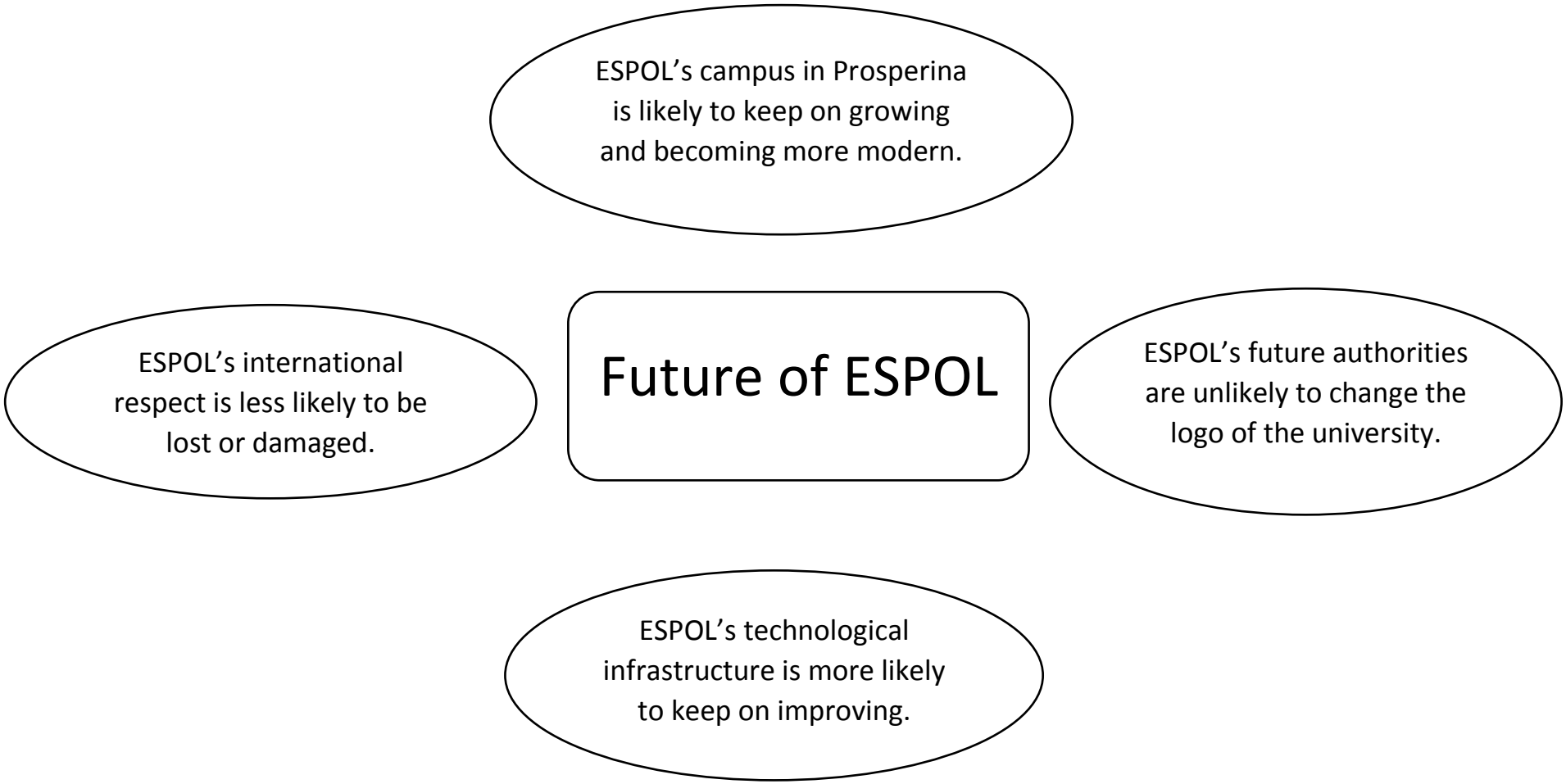
Unit 2

Here are some **predictions** related to **ESPOL's conditions in** the year **2058**. Decide **which two** of these **statements are likely to happen** and **why**.



Unit 2

Here are some **predictions** related to **ESPOL's conditions in** the year **2058**. Decide **which two** of these **statements are likely to happen** and **why**.



APPENDIX 6 – Analytic Rubric for Oral Exam Proposal

Intermediate B - Oral Exam Rubric

4. Students must take the oral exam in pairs or groups.
5. The tasks must include at least one activity where the students interact with each other without intervention by the examiner.
6. The tasks must assess the following learning outcomes:
   - ❖ Students can enter unprepared into conversations on topics that are familiar, of personal interest or relevant to their everyday life. They can give simple reasons and explanations for their opinions and plans.
   - ❖ Students can narrate their own personal experiences, and describe reactions and feelings with some prior preparation. (production)

Note: The topics that are chosen for the mid-term and final oral exams *respectively* should be based on what has been taught in that half of the course and the overall learning outcomes.

	0	0.5	1	1.5	2	2.5
Grammar and Vocabulary	The student does not appropriately use structures or vocabulary items and makes many impeding errors.	The student appropriately uses structures and vocabulary with some non-impeding errors, but the language is significantly more basic than the level of the course.	Errors with basic language items are common. The student attempts to use a limited range of structures and vocabulary from the course, but makes some impeding errors.	Errors with basic language items are regular. The student appropriately uses a limited range of structures and vocabulary from the course with some non-impeding errors.	Errors with basic language items are sporadic. The student attempts to use a wide range of structures and vocabulary from the course, but makes a few impeding errors.	Errors with basic language items are rare. The student appropriately uses a range of structures and vocabulary from the course with some non-impeding errors.
Fluency	The student cannot form more than one phrase or sentence without long periods of hesitation.	The student can form a small number of phrases or sentences after repeated periods of hesitation.	The student can form sentences together and react to most questions, but hesitation often	The student can form several sentences together and react to most questions, but occasionally hesitates	The student can form several sentences together and react to most questions, but commonly hesitates	The student can form several sentences together and react to questions without any excessive hesitation.

			impedes conversation.	for an unnatural length of time.	for a conventional length of time.	
Pronunciation	The student cannot be understood at all.	The student can be barely understood.	The student can sometimes be understood and sometimes not.	The speech sounds very unnatural. The student can be understood on the whole, but certain words are difficult to understand.	The speech sounds slightly unnatural. The student can be understood without serious effort and hardly any words are difficult to understand.	The student can be easily understood and shows fairly natural speech (intonation, stress and linking).
Interactive communication	The student cannot interact with his/her partner.	The student can respond, but cannot initiate interaction.	The student uses English to solve communication problems. The student can respond and rarely initiates, but does not develop interaction.	The student uses English to solve communication problems. The student can respond and occasionally initiates, but does not develop interaction.	The student uses English to solve communication problems. The student can respond, regularly initiates and attempts to develop interaction.	The student uses English to solve communication problems. The student can respond, spontaneously initiates and naturally develops interaction.

Four Sub-skills Criteria:** ***To provide clear and useful feedback after oral exam tasks.

	0	0.5	1	1.5	2	2.5
Grammar and Vocabulary	The student does not appropriately use structures or vocabulary items and makes many impeding errors.	The student appropriately uses structures and vocabulary with some non-impeding errors, but the language is significantly more basic than the level of the course.	Errors with basic language items are common. The student attempts to use a limited range of structures and vocabulary from the course, but makes some impeding errors.	Errors with basic language items are regular. The student appropriately uses a limited range of structures and vocabulary from the course with some non-impeding errors.	Errors with basic language items are sporadic. The student attempts to use a wide range of structures and vocabulary from the course, but makes a few impeding errors.	Errors with basic language items are rare. The student appropriately uses a range of structures and vocabulary from the course with some non-impeding errors.
Fluency	The student cannot form more than one phrase or sentence without long periods of hesitation.	The student can form a small number of phrases or sentences after repeated periods of hesitation.	The student can form sentences together and react to most questions, but hesitation often impedes conversation.	The student can form several sentences together and react to most questions, but occasionally hesitates for an unnatural length of time.	The student can form several sentences together and react to most questions, but commonly hesitates for a conventional length of time.	The student can form several sentences together and react to questions without any excessive hesitation.
Pronunciation	The student cannot be understood at all.	The student can be barely understood.	The student can sometimes be understood and sometimes not.	The speech sounds very unnatural. The student can be understood on the whole, but certain words are difficult to understand.	The speech sounds slightly unnatural. The student can be understood without serious effort and hardly any words are difficult to understand.	The student can be easily understood and shows fairly natural speech (intonation, stress and linking).

Interactive communication	The student cannot interact with his/her partner.	The student can respond, but cannot initiate interaction.	The student uses English to solve communication problems. The student can respond and rarely initiates, but does not develop interaction.	The student uses English to solve communication problems. The student can respond and occasionally initiates, but does not develop interaction.	The student uses English to solve communication problems. The student can respond, regularly initiates and attempts to develop interaction.	The student uses English to solve communication problems. The student can respond, spontaneously initiates and naturally develops interaction.
Global achievement	Very bad for their level.	Unsatisfactory for their level.	Weak for their level.	Satisfactory for their level.	Good for their level.	Excellent for their level.

Five Sub-skills Criteria:** ***For giving adequate and immediate feedback during class sessions.

APPENDIX 7 – Sample of Validation of the Structured Questionnaire in English

Carlos Daniel Cazco
Juan Carlos Delgado
ESCUELA POLITECNICA DEL LITORAL

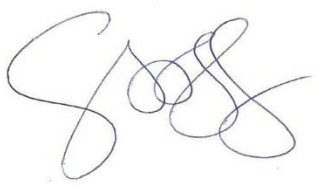
6 February 2017

Dear Daniel and Juan Carlos,

Upon your request, and drawing on my experience as an English teacher and oral examiner who uses theory to research and reflect my own practice, I was glad to analyse the questions in the questionnaire which you propose to use in your thesis study ***"Standardization of an Oral Progressive English Exam for EFL students of Pre-Intermediate level in an Ecuadorian state university: Its impact and Learners Perceptions.". I noted that these*** were adapted from an original study called ***"Attitudes and Perceptions of the Students and Instructors towards Testing Speaking Communicatively"*** made by Turan Paker Devrim Höl, at Pamukkale University and read Mr Devrim's study.

I consider that all of the questions which you have selected are relevant to your research objectives and are comprehensive and appropriately worded. I believe that applying these questions through a survey method will enable you to gain valuable insights which will be of benefit to our teaching context. I look forward to reading more about your findings.

Yours sincerely,



Graham Stagg,
MA Applied Linguistics, University of Birmingham, UK

APPENDIX 8 – Letter for Validation of Activities of Current Oral Exam Format

Esteemed colleague educator

The main objective of this letter is to ask you for a helping hand with our dissertation. We are a pair of *EFL* (English as a Foreign Language) instructors who are working on their thesis to get the degree of Master in Teaching English as a Foreign Language – *MTEFL*. You have been chosen for this task because we are aware that you are both an accomplished and an experienced professional on the field of education.

Attached to this letter, you will find these four brief documents:

1) **Appendix 1:** A sample of the current format that is being used at our present place of work (an Ecuadorian state university) to grade the speaking skill of learners whose English proficiency level is "Pre-Intermediate" or "B1", according to *CEFR* – Common European Framework of Reference for Languages: Learning, Teaching, Assessment.

2) **Appendix 2:** The analytic rubrics that are being used to grade those learners' performances following the provided sample of this institution's current oral exam.

3) **Appendix 3:** A table describing the different features comprised in each topic (i.e. language function(s) related to the topic, task type, task format, task focus and timing) that will be applied during the existing oral speaking assessment.

4) **Appendix 4:** A set of photocopies that has the actual activities that will be used by the students to develop each topic.

We need you to give us your expert and honest opinion as responses for these two questions:

1) Do those topics and activities previously mentioned (Appendixes 3 & 4) comply with the requirements that this institution's current oral exam ask for?

2) Are those topics and activities previously mentioned (Appendixes 3 & 4) adequate and suitable for this institution's current oral exam format?

Feel free to write your comments on the blank piece of paper called EXPERIENCED LANGUAGE INSTRUCTOR COMMENTS (**Appendix 5**) available at the end of this formal written request.

We truly appreciate your assistance since your unbiased feedback is going to be determinant, useful and valuable for and during the development of our dissertation.

Best regards,
Daniel Cazco
Juan Carlos Delgado Villena

EXPERIENCED LANGUAGE INSTRUCTOR COMMENTS
_____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____ _____
Signature _____ **Name** _____ **ID Number** _____ **Date** _____

*** Reference**

CEFR – Common European Framework of Reference for Languages: Learning, Teaching, Assessment. http://www.coe.int/t/dg4/linguistic/cadre1_en.asp

APPENDIX 9 – Sample of Validation for Activities of Current Oral Exam Format

EXPERIENCED LANGUAGE INSTRUCTOR COMMENTS

General comments:

The information about the current oral exam format is clear enough to offer the reader a clear picture of how the institution asses students´ oral abilities. The familiarity of the topics helps students feel comfortable during of the assessment. However, the document should include the course objectives regarding speaking. This way, the reader can find the relationship between what students do and what they are expected to do at the end of the course.

Comments per option:

OPTION 1
1. According to the rubric, yes. This exercise gives the students, the opportunity to use different structures and a lot of vocabulary according to the description/opinion.
2. a. This exercise seems to be too easy for an oral exam, since teachers might have used the same exercise in class. b. Students should be given the words but also the instruction of explaining why they predict certain things. Of course, similar exercises should be done in class, but never the same. c. The TASK FOCUS is not related to the function. Here, students have to make predictions, and express their opinions, but they do not have to mention things such as likes and habits.

OPTION 2
1. According to the rubric, yes. This exercise gives the students, the opportunity to use different structures and some vocabulary related to the description. The topic has been practiced in class. However, it does not give the students the chance to describe other objects, since they tend to talk about the same favourite possession. So, they can memorize the task.
2. a. This exercise seems to be too simple for an oral exam, since teachers might have used the same exercise in class, and students might memorize the description. b. Students should be given the words but also the instruction of explaining why they predict certain things. Of course, similar exercises should be done in class, but never the same. c. The exercise should suggest interaction between students. This one seems to be a monologue.

OPTION 3
1. According to the rubric, yes. This exercise gives the students, the opportunity to use different structures and vocabulary related to the conversation. In addition, teacher can assign different roles, like ex school classmates, colleagues, friends, neighbors, and so on. There is a lot of interaction and students will not have the possibility to memorize what they´ll say.
2. Yes, there is a lot of interaction, the time is perfect. Students can make a variety of questions, similar to the ones they have practiced in class.

OPTION 4
1. According to the rubric, yes. This exercise gives the students, the opportunity to use different structures and vocabulary related to the plan. The teacher can assign different roles, settings and others. There is a lot of interaction and memorization is not that likely to happen.
2. Yes, there is a lot of interaction, the time is perfect. Students can make a variety of questions and make adjustments to the plan.

OPTION 5
1. According to the rubric, yes. Students can use different structures and vocabulary to describe their homes and their ideal homes. There could be interaction if students have practiced questions related to the topic. However, it might happen that some students memorize the description.
2. Yes, there is a lot of interaction, the time is perfect. According to the institutional goals, the structures practiced in this lesson can be used to describe other ideal things, such as jobs, family and so on.

Signature: _____

Name: Karina León Davila.

ID number: 0914742010

Date: November 16th, 2016

APPENDIX 10 – Sample of Validation for Rubrics of the Oral Exam Proposal

Carlos Daniel Cazco
Juan Carlos Delgado
ESCUELA POLITECNICA DEL LITORAL

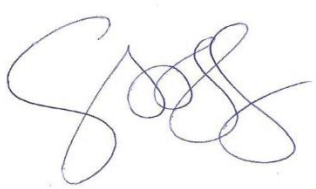
8 February 2017

Dear Daniel and Juan Carlos,

Upon your request, and drawing on my experience as an English teacher and oral examiner who uses theory to research and reflect my own practice, I was glad to analyse the rubrics which you intend to use to evaluate oral performance in class and in an exam. I consider that the rubrics are well-worded, clear, easy to understand for a competent English user (some terms might need explaining to students) and will serve the purpose of grading the students performances, in a proper manner.

I consider it is more appropriate to use the FOUR SUB-SKILLS CRITERIA for the exam because it might be difficult to apply the fifth global achievement grade independently of the other criteria.

Yours sincerely,



Graham Stagg,
MA Applied Linguistics, University of Birmingham, UK

APPENDIX 11 – Explanation of formulae for Student's *t* test

The statistical processes compulsory for this analytics trial demand the use of certain specific formulae to calculate the fundamental measures of standard deviation, arithmetic mean, and standard error for the media.

For *Standard Deviation*:

$$\boldsymbol{S} = \sqrt{\frac{\sum D^2}{N} - (\overline{X_1} - \overline{X_2})^2}$$

Wherein:

S = Standard deviation of the distribution of the differences in scores before applying the intervention to the students and after employing it, considering the deviations of the scores with respect to their arithmetic mean.

D = Difference between the score obtained before the intervention and the score obtained after it.

N = The number of students evaluated in the sample.

For the *Arithmetic Mean*:

$$\overline{X_1} = \frac{\sum X_1}{N} \qquad \overline{X_2} = \frac{\sum X_2}{N}$$

Wherein:

$\overline{X_1}$ = Average of the first measurement.

$\overline{X_2}$ = Average of the second measurement.

$\sum X_1$ = Sum of the first measurements.

$\sum X_2$ = Sum of the second measurements.

N = The number of students evaluated in the sample.

For the *standard error for the Media*:

$$\boldsymbol{O_{dif}} = \frac{S}{\sqrt{\mathrm{N}-1}}$$

O dif = Standard error of the difference.

S = Standard deviation of the distribution of the differences in scores before applying the intervention to the students and after employing it, considering the deviations of the scores with respect to their arithmetic mean.

N = The number of students evaluated in the sample.

To draw this segment of formulae for mathematical measurements to a close, here is the formula for the *Test of Statistical Significance*:

$$\mathbf{t} = \frac{\overline{\mathrm{X}_1} - \overline{\mathrm{X}_2}}{\mathrm{O}_{\mathrm{dif}}}$$

t = Statistical significance test that establishes the difference between sample means before applying the intervention to the students and after employing it.

APPENDIX 12 – Student's *t* test of Group "A"

Next, the reader will find the oral exams scores from the collection of students under the training of "RESEARCHER A" or "GROUP A" from now on.

Step 1: The process matrix is constructed respecting the data obtained in the research.

Table 4.3 Scores of Entry Exams (2016) and Exit Exams (2017) for Group "A"

GROUP "A" – SCORES OF ORAL EXAMS				
#	ENTRY EXAM	EXIT EXAM	D	D^2
	X1	X2	X1 - X2	$(X1 - X2)^2$
1	6,70	9,5	-2,80	7,84
2	7,30	10,00	-2,70	7,29
3	4,70	10,00	-5,30	28,09
4	5,30	9,70	-4,40	19,36
5	7,30	10,00	-2,70	7,29
6	3,30	8,70	-5,40	29,16
7	6,70	10,00	-3,30	10,89
8	7,30	10,00	-2,70	7,29
9	8,70	9,00	-0,30	0,09
10	3,30	8,00	-4,70	22,09
11	6,00	7,50	-1,50	2,25
12	4,70	9,50	-4,80	23,04
13	6,00	9,70	-3,70	13,69
14	4,70	9,20	-4,50	20,25
15	4,70	9,20	-4,50	20,25
16	2,00	9,20	-7,20	51,84
17	8,00	10,00	-2,00	4,00
18	8,00	10,00	-2,00	4,00
19	8,70	10,00	-1,30	1,69
20	4,70	9,00	-4,30	18,49
21	5,30	9,50	-4,20	17,64
22	5,30	9,50	-4,20	17,64
23	4,70	8,00	-3,30	10,89
24	3,30	5,00	-1,70	2,89
25	8,70	10,00	-1,30	1,69
26	3,30	9,00	-5,70	32,49
27	7,30	8,60	-1,30	1,69
28	3,30	3,00	0,30	0,09

29	4,00	10,00	-6,00	36,00
30	3,30	8,50	-5,20	27,04
31	3,30	8,20	-4,90	24,01
32	9,30	10,00	-0,70	0,49
33	6,70	10,00	-3,30	10,89
34	6,70	9,20	-2,50	6,25
35	9,30	9,20	0,10	0,01
∑	201,90	315,90	-114,00	488,60
Average	5,77	9,03		

In the specific case of "GROUP A", the results from the entry evaluations to the exit examinations varied positively from lower numbers to higher figures on the whole. This change means, in general terms, that these students performed better on this oral exam proposal than on the current format for the speaking test used by this language center.

Step 2.1: Find the Arithmetic Media required by the Student's t-test formula. These numbers will be calculated by considering the results from the entry and exit oral examinations taken by these learners.

$$\overline{X_1} = \frac{\sum X_1}{N} \qquad \overline{X_2} = \frac{\sum X_2}{N}$$

$$\overline{X_1} = \frac{201,90}{35} \qquad \overline{X_2} = \frac{315,90}{35}$$

$$\boldsymbol{\overline{X_1}} = 5,77 \qquad \boldsymbol{\overline{X_2}} = 9,03$$

Step 2.2: Find the Standard Deviation required by the formula. The figures to be used will come from the results gotten by these students in these entry and exit spoken tests.

$$\boldsymbol{S} = \sqrt{\frac{\sum D^2}{N} - (\overline{X_1} - \overline{X_2})^2}$$

$$\boldsymbol{S} = \sqrt{\frac{488,60}{35} - (5,77 - 9,03)^2}$$

$$\boldsymbol{S} = 1.83$$

Step 2.3: Find the Standard Error of the Difference.

$$\boldsymbol{O_{dif}} = \frac{\mathrm{S}}{\sqrt{\mathrm{N}-1}}$$

$$\boldsymbol{O_{dif}} = \frac{1{,}83}{\sqrt{35-1}}$$

$$\boldsymbol{O_{dif}} = 0{,}31$$

Step 3: The value of the statistical significant test known as student's t-test (*t*) is calculated.

Table 4.4 Values of the required measures for calculating the value of *t* for Group "A"

MEASURES	VALUES	
	ENTRY	EXIT
Arithmetic Media	5,77	9,03
Standard Deviation	1,83	
Standard Error	0,31	
N	35	

$$\mathbf{t} = \frac{\overline{\mathrm{X_1}} - \overline{\mathrm{X_2}}}{\mathrm{O_{dif}}}$$

$$\mathbf{t} = \frac{5{,}77 - 9{,}03}{0{,}31}$$

$$\mathbf{t} = \mathbf{-10{,}52}$$

Step 4: The analysis of the value of *t* is carried out.

Data Analysis and Interpretation: The *t* value calculated is *-10,52*. This value is well outside the boundaries from the region of *not acceptance* after comparing it to the critical limits, which according to the 0.05 significance level or 0.975 critical value for 35 participants are any values that are greater than ± 2,030 in the Student's *t* Distribution Table. (NIST/SEMATECH, 2013) This numerical outcome clearly reflects the complete acceptance of the new verbal assessment scheme by these students.

Step 5: A decision is made based upon the interpretation of *t*, which basically means that the result obtained is contrasted against the null hypothesis for its acceptance or rejection.

Decision: The end results these learners got in their exit oral exam (i.e. oral exam proposal) were much better than the ones obtained in their entry spoken test thus the null hypothesis for this set of undergraduates was rejected. Therefore, "RESEARCHER A" could affirm that his students fully approved this new verbal assessment scheme as well as the strategies used throughout this intervention because his pupils showed competent domain of them in the process of this vocal examination.

APPENDIX 13 – Student's *t* test of Group "B"

The chart below contains the grades that the students from "GROUP B" received on their entry and exit spoken tests, which are the numbers that will be explored in the subsequent lines.

Step 1: The process matrix is constructed respecting the data obtained in the research.

Table 4.5 Scores of Entry Exams (2016) and Exit Exams (2017) for Group "B"

GROUP "B" – SCORES OF ORAL EXAMS				
#		EXIT EXAM	D	D^2
		X2	X1 - X2	$(X1 - X2)^2$
1	10,00	4,25	5,75	33,06
2	10,00	2,00	8,00	64,00
3	0,00	5,25	-5,25	27,56
4	9,00	7,75	1,25	1,56
5	10,00	9,50	0,50	0,25
6	10,00	6,25	3,75	14,06
7	10,00	4,75	5,25	27,56
8	9,00	4,00	5,00	25,00
9	10,00	7,25	2,75	7,56
10	10,00	2,50	7,50	56,25
11	9,00	4,25	4,75	22,56
12	10,00	9,00	1,00	1,00
13	8,00	2,25	5,75	33,06
14	9,00	2,25	6,75	45,56
15	10,00	3,25	6,75	45,56
16	10,00	5,00	5,00	25,00
17	10,00	5,50	4,50	20,25
18	10,00	7,50	2,50	6,25
19	9,00	5,00	4,00	16,00
20	10,00	6,75	3,25	10,56
21	10,00	9,50	0,50	0,25
22	9,00	4,00	5,00	25,00
23	10,00	5,50	4,50	20,25
24	9,00	5,75	3,25	10,56
25	9,00	2,75	6,25	39,06
26	9,00	9,00	0,00	0,00

27	10,00	4,25	5,75	33,06
28	10,00	5,75	4,25	18,06
29	9,00	4,50	4,50	20,25
30	10,00	6,50	3,50	12,25
∑	278,00	161,75	116,25	661,44
Average	9,27	5,39		

The table above shows that the results of the entry and exit examinations for the constituents of "GROUP B" differed from higher to lower quantities and that is the main reason why these researchers decided to examine these figures separately from those numbers belonging to "GROUP A". If this information set (i.e. "GROUP B") had been analyzed together with the figures from the other assembly (i.e. "GROUP A") then the general statistics of the study would have been evidently affected, and what is worst in an extremely negative way.

Step 2.1: Find the Arithmetic Media required by the Student's t-test formula. These numbers will be calculated by considering the results from the entry and exit oral examinations taken by these learners.

$$\overline{X_1} = \frac{\sum X_1}{N} \qquad \overline{X_2} = \frac{\sum X_2}{N}$$

$$\overline{X_1} = \frac{278{,}00}{30} \qquad \overline{X_2} = \frac{161{,}75}{30}$$

$$\boldsymbol{\overline{X_1}} = 9{,}27 \qquad \boldsymbol{\overline{X_2}} = 5{,}39$$

Step 2.2: Find the Standard Deviation required by the formula. The figures to be used will come from the results gotten by these students in these entry and exit spoken tests.

$$\boldsymbol{S} = \sqrt{\frac{\sum D^2}{N} - (\overline{X_1} - \overline{X_2})^2}$$

$$\boldsymbol{S} = \sqrt{\frac{661{,}44}{30} - (9{,}27 - 5{,}39)^2}$$

$$\boldsymbol{S} = 2{,}64$$

Step 2.3: Find the Standard Error of the Difference.

$$\boldsymbol{O_{dif}} = \frac{\mathrm{S}}{\sqrt{\mathrm{N}-1}}$$

$$\boldsymbol{O_{dif}} = \frac{2{,}64}{\sqrt{30-1}}$$

$$\boldsymbol{O_{dif}} = 0{,}49$$

Step 3: The value of the statistical significant test known as student's t-test (*t*) is calculated.

Table 4.6 Values of the required measures for calculating the value of *t* for Group "B"

MEASURE	VALUE	
	ENTRY	EXIT
Arithmetic Media	9,27	5,39
Standard Deviation	2,64	
Standard Error	0,49	
N	30	

$$\mathbf{t} = \frac{\overline{\mathrm{X_1}} - \overline{\mathrm{X_2}}}{\mathrm{O_{dif}}}$$

$$\mathbf{t} = \frac{9{,}27 - 5{,}39}{0{,}49}$$

$$\mathbf{t = 7,91}$$

Step 4: The analysis of the value of *t* is carried out.

Data Analysis and Interpretation: The *t* value calculated is *7,91* because the end results gotten by these pupils in their exit oral exam (i.e. new verbal assessment scheme) were different or less – as a general rule – than the first scores obtained in their entry oral exam. This value once again is well outside the boundaries from the region of *not acceptance* after comparing it to the critical limits, which according to the 0.05 significance level or 0.975 critical value for 30 participants are any values that are greater than ±2,042 in the Student's *t* Distribution Table. (NIST/SEMATECH, 2013) However, this statistical product evidently mirrors that the students of this specific

collection had stern problems during the process of this oral exam proposal, especially in the “interactive turn”.

Step 5: A decision is made based upon the interpretation of *t*, which basically means that the result obtained is contrasted against the null hypothesis for its acceptance or rejection.

Decision: The scores obtained by these students after doing this new verbal assessment scheme were dissimilar and lower compared to the marks of the current oral exam consequently the null hypothesis for this set of undergraduates was rejected as well. Except this rejection does not signify that there was a significantly positive change in the sample; quite the opposite, it means that these students had serious difficulties during this innovative spoken test hence the strategies exercised for the EFL instructor in charge of their guidance during this educational project were not quite suitable for the linguistics needs of these particular learners.

APPENDIX 14 – Table with Critical Values of the Student's *t* Distribution

This information about ***Critical Values of the Student's t Distribution*** was retrieved from https://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm.

1.3.6.7.2. Critical Values of the Student's *t* Distribution

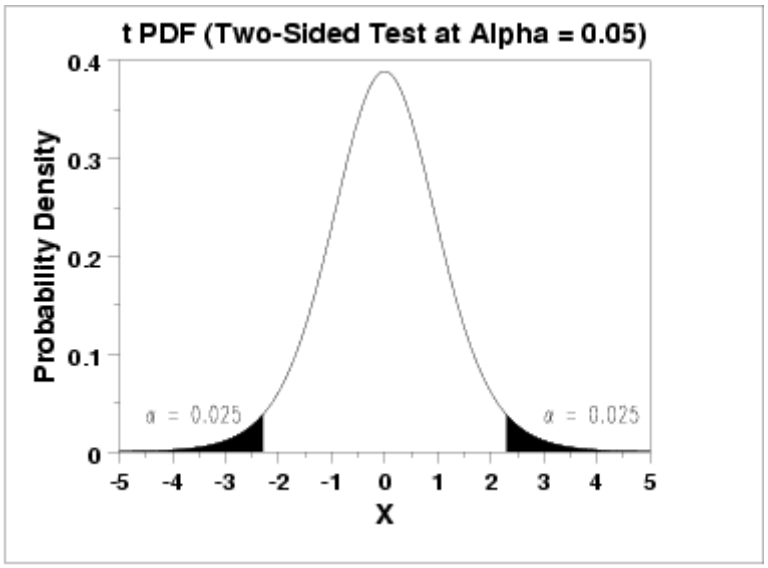
How to Use This Table

This table contains critical values of the Student's t distribution computed using the cumulative distribution function. The t distribution is symmetric so that

$$t_{1-\alpha,\nu} = -t_{\alpha,\nu}.$$

The t table can be used for both one-sided (lower and upper) and two-sided tests using the appropriate value of α.

The significance level, α, is demonstrated in the graph below, which displays a t distribution with 10 degrees of freedom. The most commonly used significance level is $\alpha = 0.05$. For a two-sided test, we compute $1 - \alpha/2$, or $1 - 0.05/2 = 0.975$ when $\alpha = 0.05$. If the absolute value of the test statistic is greater than the critical value (0.975), then we reject the null hypothesis. Due to the symmetry of the t distribution, we only tabulate the positive critical values in the table below.



Given a specified value for α :

1. For a two-sided test, find the column corresponding to 1-α/2 and reject the null hypothesis if the absolute value of the test statistic is greater than the value of $t_{1-\alpha/2,\,\nu}$ in the table below.
2. For an upper, one-sided test, find the column corresponding to 1-α and reject the null hypothesis if the test statistic is greater than the table value.
3. For a lower, one-sided test, find the column corresponding to 1-α and reject the null hypothesis if the test statistic is less than the negative of the table value.

Critical values of Student's *t* distribution with *v* degrees of freedom

Probability less than the critical value ($t_{1-\alpha,\,\nu}$)

ν	0.90	0.95	0.975	0.99	0.995	0.999
1.	3.078	6.314	12.706	31.821	63.657	318.313
2.	1.886	2.920	4.303	6.965	9.925	22.327
3.	1.638	2.353	3.182	4.541	5.841	10.215
4.	1.533	2.132	2.776	3.747	4.604	7.173
5.	1.476	2.015	2.571	3.365	4.032	5.893
6.	1.440	1.943	2.447	3.143	3.707	5.208
7.	1.415	1.895	2.365	2.998	3.499	4.782
8.	1.397	1.860	2.306	2.896	3.355	4.499
9.	1.383	1.833	2.262	2.821	3.250	4.296
10.	1.372	1.812	2.228	2.764	3.169	4.143
11.	1.363	1.796	2.201	2.718	3.106	4.024
12.	1.356	1.782	2.179	2.681	3.055	3.929
13.	1.350	1.771	2.160	2.650	3.012	3.852
14.	1.345	1.761	2.145	2.624	2.977	3.787
15.	1.341	1.753	2.131	2.602	2.947	3.733
16.	1.337	1.746	2.120	2.583	2.921	3.686
17.	1.333	1.740	2.110	2.567	2.898	3.646
18.	1.330	1.734	2.101	2.552	2.878	3.610
19.	1.328	1.729	2.093	2.539	2.861	3.579
20.	1.325	1.725	2.086	2.528	2.845	3.552
21.	1.323	1.721	2.080	2.518	2.831	3.527
22.	1.321	1.717	2.074	2.508	2.819	3.505
23.	1.319	1.714	2.069	2.500	2.807	3.485
24.	1.318	1.711	2.064	2.492	2.797	3.467
25.	1.316	1.708	2.060	2.485	2.787	3.450
26.	1.315	1.706	2.056	2.479	2.779	3.435
27.	1.314	1.703	2.052	2.473	2.771	3.421
28.	1.313	1.701	2.048	2.467	2.763	3.408
29.	1.311	1.699	2.045	2.462	2.756	3.396
30.	1.310	1.697	2.042	2.457	2.750	3.385
31.	1.309	1.696	2.040	2.453	2.744	3.375
32.	1.309	1.694	2.037	2.449	2.738	3.365
33.	1.308	1.692	2.035	2.445	2.733	3.356
34.	1.307	1.691	2.032	2.441	2.728	3.348
35.	1.306	1.690	2.030	2.438	2.724	3.340
36.	1.306	1.688	2.028	2.434	2.719	3.333
37.	1.305	1.687	2.026	2.431	2.715	3.326
38.	1.304	1.686	2.024	2.429	2.712	3.319
39.	1.304	1.685	2.023	2.426	2.708	3.313
40.	1.303	1.684	2.021	2.423	2.704	3.307
41.	1.303	1.683	2.020	2.421	2.701	3.301
42.	1.302	1.682	2.018	2.418	2.698	3.296

43.	1.302	1.681	2.017	2.416	2.695	3.291
44.	1.301	1.680	2.015	2.414	2.692	3.286
45.	1.301	1.679	2.014	2.412	2.690	3.281
46.	1.300	1.679	2.013	2.410	2.687	3.277
47.	1.300	1.678	2.012	2.408	2.685	3.273
48.	1.299	1.677	2.011	2.407	2.682	3.269
49.	1.299	1.677	2.010	2.405	2.680	3.265
50.	1.299	1.676	2.009	2.403	2.678	3.261
51.	1.298	1.675	2.008	2.402	2.676	3.258
52.	1.298	1.675	2.007	2.400	2.674	3.255
53.	1.298	1.674	2.006	2.399	2.672	3.251
54.	1.297	1.674	2.005	2.397	2.670	3.248
55.	1.297	1.673	2.004	2.396	2.668	3.245
56.	1.297	1.673	2.003	2.395	2.667	3.242
57.	1.297	1.672	2.002	2.394	2.665	3.239
58.	1.296	1.672	2.002	2.392	2.663	3.237
59.	1.296	1.671	2.001	2.391	2.662	3.234
60.	1.296	1.671	2.000	2.390	2.660	3.232
61.	1.296	1.670	2.000	2.389	2.659	3.229
62.	1.295	1.670	1.999	2.388	2.657	3.227
63.	1.295	1.669	1.998	2.387	2.656	3.225
64.	1.295	1.669	1.998	2.386	2.655	3.223
65.	1.295	1.669	1.997	2.385	2.654	3.220
66.	1.295	1.668	1.997	2.384	2.652	3.218
67.	1.294	1.668	1.996	2.383	2.651	3.216
68.	1.294	1.668	1.995	2.382	2.650	3.214
69.	1.294	1.667	1.995	2.382	2.649	3.213
70.	1.294	1.667	1.994	2.381	2.648	3.211
71.	1.294	1.667	1.994	2.380	2.647	3.209
72.	1.293	1.666	1.993	2.379	2.646	3.207
73.	1.293	1.666	1.993	2.379	2.645	3.206
74.	1.293	1.666	1.993	2.378	2.644	3.204
75.	1.293	1.665	1.992	2.377	2.643	3.202
76.	1.293	1.665	1.992	2.376	2.642	3.201
77.	1.293	1.665	1.991	2.376	2.641	3.199
78.	1.292	1.665	1.991	2.375	2.640	3.198
79.	1.292	1.664	1.990	2.374	2.640	3.197
80.	1.292	1.664	1.990	2.374	2.639	3.195
81.	1.292	1.664	1.990	2.373	2.638	3.194
82.	1.292	1.664	1.989	2.373	2.637	3.193
83.	1.292	1.663	1.989	2.372	2.636	3.191
84.	1.292	1.663	1.989	2.372	2.636	3.190
85.	1.292	1.663	1.988	2.371	2.635	3.189
86.	1.291	1.663	1.988	2.370	2.634	3.188
87.	1.291	1.663	1.988	2.370	2.634	3.187
88.	1.291	1.662	1.987	2.369	2.633	3.185
89.	1.291	1.662	1.987	2.369	2.632	3.184
90.	1.291	1.662	1.987	2.368	2.632	3.183
91.	1.291	1.662	1.986	2.368	2.631	3.182
92.	1.291	1.662	1.986	2.368	2.630	3.181
93.	1.291	1.661	1.986	2.367	2.630	3.180
94.	1.291	1.661	1.986	2.367	2.629	3.179
95.	1.291	1.661	1.985	2.366	2.629	3.178
96.	1.290	1.661	1.985	2.366	2.628	3.177
97.	1.290	1.661	1.985	2.365	2.627	3.176
98.	1.290	1.661	1.984	2.365	2.627	3.175
99.	1.290	1.660	1.984	2.365	2.626	3.175
100.	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.090