



THE TESTING OF LISTENING IN BILINGUAL SECONDARY SCHOOLS OF COSTA RICA: BRIDGING GAPS BETWEEN THEORY AND PRACTICE

EVALUACIÓN DE LA DESTREZA AUDITIVA EN LOS COLEGIOS BILINGÜES DE COSTA
RICA: ACORTANDO BRECHAS ENTRE TEORÍA Y PRÁCTICA

Volumen 14, Número 2

Mayo - Agosto
pp. 1-23

Este número se publicó el 30 de mayo de 2014

Roy Gamboa Mena
Henry Sevilla Morales

Revista indizada en [REDALYC](#), [SCIELO](#)

Revista distribuida en las bases de datos:

[CATÁLOGO DE LATINDEX](#), [IRESIE](#), [CLASE](#), [DIALNET](#), [DOAJ](#), [E-REVIST@S](#),
[SHERPA/ROMEO](#), [QUALIS](#), [MIAR](#)

Revista registrada en los directorios:

[ULRICH'S](#), [REDIE](#), [RINACE](#), [OEI](#), [MAESTROTECA](#), [PREAL](#), [CLASCO](#)

Los contenidos de este artículo están bajo una licencia [Creative Commons](#)



THE TESTING OF LISTENING IN BILINGUAL SECONDARY SCHOOLS OF COSTA RICA: BRIDGING GAPS BETWEEN THEORY AND PRACTICE

EVALUACIÓN DE LA DESTREZA AUDITIVA EN LOS COLEGIOS BILINGÜES DE COSTA RICA: ACORTANDO BRECHAS ENTRE TEORÍA Y PRÁCTICA

Roy Gamboa Mena¹
Henry Sevilla Morales²

Abstract: *Despite listening being one of the most crucial skills in the process of communication, research shows that it has been neglected in most English as a Foreign Language (EFL) programs, both worldwide and in Costa Rica. Worse yet, mismatches between theory and practice often result in poor listening assessment in many institutions. Thus, this article examined current listening testing practices by Ministry of Public Education (in Spanish, MEP) in-service teachers ranked C1 according to the Common European Framework (CEF) in bilingual secondary schools of the West Area of Costa Rica. Listening tests created and administered by those teachers were analyzed for their compliance with both theory on listening assessment and MEP's guidelines on assessment. The study revealed that even though teachers had previously received training on testing, the tests they created do not fully comply with both MEP's guidelines and theoretical principles on listening assessment. Findings expand conclusions drawn by Gamboa and Sevilla (2013) in previous research on listening assessment and provide several contributions to the current bulk of literature on listening testing practices in Costa Rica. Such conclusions also reveal areas of listening assessment that need to be further tackled through teacher training.*

Keywords: LISTENING SKILLS, ENGLISH, TEACHING ENGLISH, TEACHERS, BILINGUAL HIGH SCHOOLS, WEST AREA, MINISTRY OF PUBLIC EDUCATION, COSTA RICA

Resumen: *Pese a su importancia en el proceso de la comunicación, la habilidad auditiva ha sido ignorada en muchos programas de inglés como lengua extranjera (en inglés, EFL). Peor aún, discrepancias entre teoría y práctica conllevan a un inadecuado proceso de evaluación de esta habilidad lingüística. Ante ello, el presente estudio examina las prácticas en la evaluación auditiva de docentes de inglés en servicio del Ministerio de Educación Pública (MEP) ubicados en la banda C1 según el Marco Común Europeo (en inglés, CEF) en colegios bilingües de la Región de Occidente de Costa Rica. Se analizaron pruebas de escucha diseñadas por el personal docente participante en la investigación en términos de su cumplimiento con la teoría sobre evaluación auditiva y los lineamientos de evaluación del MEP. El estudio reveló que, a pesar de que los docentes habían recibido capacitación en materia de evaluación auditiva, las pruebas que ellos diseñaron no cumplen a cabalidad con los lineamientos de evaluación del MEP ni con los principios teóricos en evaluación auditiva. Los resultados ampliaron conclusiones postuladas por Gamboa y Sevilla (2013) en estudios anteriores sobre el tema y vigorizan el estado de la cuestión sobre la evaluación del componente auditivo en Costa Rica. Dichos resultados también dilucidaron áreas de la evaluación de las destrezas auditivas que aun deben trabajarse mediante la capacitación de docentes.*

Palabras clave: DESTREZA AUDITIVA, INGLÉS, ENSEÑANZA DEL INGLÉS, PROFESORADO, COLEGIOS BILINGÜES, REGIÓN DE OCCIDENTE, MINISTERIO DE EDUCACIÓN PÚBLICA, COSTA RICA

¹ Profesor de la Carrera de Bachillerato y Licenciatura en la Enseñanza del Inglés de la Universidad de Costa, Sede de Occidente. Magister en la Enseñanza del Inglés como Lengua Extranjera. Dirección Electrónica: gamboa.roy@gmail.com

² Profesor de la Carrera de Bachillerato y Licenciatura en la Enseñanza del Inglés de la Universidad de Costa Rica, Sede de Occidente. Licenciado en la Enseñanza del Inglés como Lengua Extranjera. Dirección electrónica: al_deron@hotmail.com

Artículo recibido: 5 de setiembre, 2013

Devuelto para corrección: 7 de mayo, 2014

Aprobado: 15 de mayo, 2014

1. Introduction

For many decades, teachers and researchers believed that listening was a passive skill because it could not be observed (Alderson and Bachman, 2001, in Buck, 2001); but for many decades, they were wrong. Such beliefs can be traced back to the audiolingual times when listening assessment was supposed to be assessed through the *discrete point* approach, where learners were assessed for their capacity to isolate language rather than to grasp its meaning in context (Coombe, Folse, and Hubley, 2007, p. 91). Today, however, second language researchers and academics agree that it is a very active skill, for the "students receive, construct meaning from, and respond to spoken messages" (Coombe et al., 2007, p. 90). They also believe that, besides comprehending the language, learners should be able to take that input and bring it to use in real-life contexts. Unfortunately, not always are these principles brought to practice in L2 teaching; and to worsen the scenario further, gaps exist between listening assessment theory and practice, as well as between "listening research and practice" (Osada, 2004, p. 57).

In the context of the Costa Rican public education system, research suggests serious mismatches between theory on language assessment, the assessment guidelines teachers are demanded to follow, and what they in fact do in their classrooms (Gamboa and Sevilla, 2013a, p. 24). Gamboa and Sevilla (2013) believe that the issue stems, in part, because concrete guidelines have yet not been provided by MEP; the only document available is the procedures for test design in general, not for language testing. Arguably, it is not surprising to find teaching scenarios where the testing of listening is conducted poorly, and more than that, often in an instinctive fashion. As a result of this, teachers face a number of limitations that range from not knowing the format type to follow in the construction of their tests to more serious issues such as poor content validity of their examinations. Directly, these limitations affect their teaching as a whole and, more critically, the learning process. If a test measures what it is not supposed to, for example, then learners are likely to perform poorly in the test. Thus, the study herein arises from the need to expand prior research on listening assessment conducted in Costa Rica.

Arguably, in an age of multilingualism, where efforts are being made to put Costa Rica at the forefront of international communication, and where English skills are paramount in reaching such a goal, "research on listening assessment proves not only relevant but also crucial as a way to provide insights on how to conduct better teaching in the context of

English as a Foreign Language" (Gamboa and Sevilla, 2013b, p. 187). In this manner, this study will enrich the existing bulk of literature on the testing of listening and will, in turn, help teachers and academics confront the challenges L2 listening testing has brought about in recent decades.

Thus, upon analyzing the importance of the topic, the question that emerges is: To what extent is training on listening assessment that includes both theory and MEP's testing guidelines correlative to effective testing practices in bilingual middle schools of the West Area of Costa Rica? In order to respond to that question, the goal of this paper is to study the extent to which training on listening assessment has an incidence on the actual testing practices of teachers ranked C1 in bilingual secondary schools of the West Area of Costa Rica. To that purpose, teacher-created tests were analyzed quantitatively for their compliance with both theory on listening assessment and MEP's guidelines on general assessment. The resulting data underwent a process of triangulation with qualitative annotations made to each of the eleven categories evaluated in the tests. Finally, the results of the tests analysis were contrasted against those of previous studies on listening assessment practices in the West Area of Costa Rica.

2. Literature Review

2.1 A Brief History of Listening Assessment

The history of listening assessment can be traced back to as early as the mid twentieth century, when experts devised two approaches to language teaching: The audiolingual and the communicative approaches (Gamboa and Sevilla, 2013b, p.187). From the audiolingual times to the advent of the communicative approaches to language teaching (from the 50s to the 80s), three approaches to listening assessment were developed: the *discrete point approach*, the *integrative approach*, and the *communicative approach*. According to Coombe et al. (2009), the first departed from the notion that in order to be able to measure a learner's mastery of the language, it was necessary to break it down into small segments or units (p. 91). Thus, typical exercises in a listening test would include phonemic discrimination or paraphrase recognition, and the learners were not expected to understand language in discourse. The second was based on the idea that "the whole of language was seen as being better than the sum of its parts" (as cited in Gamboa and Sevilla, 2013b, p. 187), and that learners had to be able to "use many bits [of language] at the same time" (Oller, 1979, p. 7).

Typical exercises using this approach included dictation and cloze. The last approach to listening assessment is found within the domains of the communicative approach, whose rationale was that language had to be comprehended in discourse and then used in a contextualized fashion (Oller, 1979, p. 7). Suggested exercises for assessment using this approach include, among many others, communicative question formats that are authentic in nature. As suggested by current theory, testing practices should be oriented by this last approach.

2.2 Methods for Assessing Listening Comprehension

In an attempt to come across ways to effectively assess listening, research has suggested a number of ways to do it. In this respect, Nunan (2002, p. 176) has suggested two approaches, *bottom-up processing* and *top-down processing*. In bottom-up processing, comprehension occurs when the listener successfully decodes the spoken text. Thus, sounds can range from the smallest meaningful units to complete texts, and comprehension occurs when students take in a word, decode it, and link it with other words to form sentences—ultimately to form meaningful texts. In top-down listening, the listener is directly involved with constructing meaning from input. The student uses background knowledge of the context and situations to make sense of what is heard. According to Nunan (2002, p. 176), the two modes are important and must therefore be taught in the class and later on assessed.

2.3 Theoretical and Practical Gaps

Despite all the bulk of emerging literature within the scope of listening comprehension and, in particular of listening assessment, research concludes that still much remains to be done to fill gaps between both theory and practice. Vandergrift (1997), for instance, referred to listening as being "the Cinderella of communication strategies" (in Gamboa and Sevilla, 2013a, p. 22). He has asserted that listening has been a neglected component in many EFL programs, and that answers must therefore be sought as to how to conduct better listening teaching practices. Along the same lines, Osada has stated that (as cited in Brown, 1987) listening in language teaching and learning is for the most part undermined. Osada (2004) has gone on to add that, despite recent awareness on its importance, listening "remains a somewhat neglected and poorly taught aspect of English..." in many ESL and EFL programs (p. 57). It follows then that, in light of Costa Rica's goal for multilingualism, these issues

deserve more attention than we generally realize. In this respect, Presidencia de la República (2007) has stated that Costa Rica's main goal regarding multilingualism is to provide the country with a population whose communicative competences enable its overall progress so that individuals have access to better education and employment.

Arguably, this somewhat big goal poses challenges that need to be met through high quality education and willpower, which have not yet been fully addressed in the context of the Costa Rican public education system. Then again, when it comes to the assessment of listening skills, in the program for III cycle in bilingual secondary schools, Presidencia de la República (2007) has dictated six "principles for assessing listening" (p. 25), but, as argued previously by Gamboa and Sevilla (2013a), these principles do not concretely orient teachers in designing their listening tests.

2.4 Recent Studies within the Context of Costa Rica's Public Education System

It is evident that little research has been done in Costa Rica regarding current in-service MEP teachers' listening assessment practices. Recently, however, there has been an increased interest in conducting research within this area. Perhaps the most recent studies in this line have been conducted by Gamboa and Sevilla (2013). Their two investigations *Assessment of Listening Skills in Public Schools of Costa Rica: The West and Central Pacific Case* (2013a), and *The Impact of Teacher Training on the Assessment of Listening Skills* (2013b) have shed some light as to what is being done in terms of assessment in the areas of San Ramón, Palmares, Alfaro Ruiz, Valverde Vega, Esparza, Puntarenas, Barranca, and other regions of the West and Central Pacific Costa Rica.

In the first study, the authors compared current listening assessment practices and beliefs of MEP teachers ranked B1 according to the CEF via analyzing listening tests created by them for their compliance with both MEP's assessment guidelines and current theory on the assessment of listening skills. The researchers found, among other results, that mismatches exist "between what teachers think they do in terms of assessment and what their actual practices are" (Gamboa & Sevilla, 2013a, p. 24) for the two areas inquired. They also concluded that "further training on the application of assessment principles is needed so that is closed between the teachers' beliefs and their current practices in terms of creation and administration of listening assessment" (Gamboa & Sevilla, 2013a, p. 25). The results

suggest that, as agreed by Vandergrift (1997) and by Osada (2004), listening is still being undermined and that more needs to be done to rescue the "Cinderella" of language teaching and learning.

In the second study, Gamboa and Sevilla (2013b) analyzed the impact that teacher training has on MEP teachers' listening assessment practices. To this end, they offered fifteen teachers ranked B1 from the west area an eight-hour workshop where listening assessment theory and MEP's assessment guidelines were studied. At the end of the workshop, these teachers created listening tests by applying the assessment principles discussed in the workshop. The researchers later analyzed the tests for compliance with both MEP's guidelines on assessment and theory on listening assessment via an adaptation of the checklist used in their first study (i.e., Assessment of Listening ...), and the results were later compared with those obtained by analyzing tests of a control group (which did not receive the training). In general, the researchers have concluded that "better listening test-design practices could be achieved by simply providing teachers with some training on listening assessment" (Gamboa & Sevilla, 2013b, p. 196). Nonetheless, they have also concluded that there are areas that need improvement such as "the writing of general instructions, specific instructions, the inclusion of general test objectives, and the improvement of listening test techniques" (Gamboa & Sevilla, 2013b, p. 196).

2.5 Gaps in Recent Research on Listening Assessment in the Context of Costa Rica

Despite these recent efforts to elucidate listening assessment practices in Costa Rica's public education system, it is evident that a dramatic gap still needs to be filled between theory and practice. Research has not yet explored the listening practices of teachers ranked C1 and who have received while-in-service training on listening assessment. Furthermore, Gamboa and Sevilla (2013) experienced some limitations in their previous studies, which included, among others, "examining the listening passages [of the tests], deal[ing] with the time constrictions experienced in [their] study", and conducting similar studies with populations ranked in different levels (Gamboa & Sevilla, 2013b, p. 196).

Another important gap to highlight here is the fact that although Gamboa and Sevilla's training workshop in their last study (i.e., 2013b) proved generally successful, no tests collected from the natural teaching setting have yet been analyzed in the West Area, namely

because test design at the workshop occurred in a controlled environment that did not entirely resemble the conditions in which tests are usually created for regular classroom use. This implies that there is a degree of likelihood for those tests to have been created as the mere act of complying with the requirements of a workshop, and that results may have been influenced by what Porte calls "THE HAWTHORNE EFFECT"; a condition in which participants "react in a way that is related to their pleasure at being included in a study rather than to any treatment involved" (Porte, 2010, p. 103). The tests collected for the present study, on the contrary, were tests previously designed and administered by C1 in-service English teachers who were not told about their participation in any study before they created them.

This section has presented the most relevant theory that comprises the backbone of the present study. Firstly, it presented a short historical account of listening assessment. Secondly, it gave a review of two models of listening proposed by Nunan (2002), highlighting the need for more research on listening assessment as suggested by research. Lastly, this section has outlined two of the most recent contributions to the field of listening assessment in the context of national education of Costa Rica. Hence, the pages that follow will deal with the methodology and the procedures that will support the development of the research project herein described.

3 Methodology

3.1 The Research Design

The present study used a mixed paradigm, namely theoretical principles of both quantitative and qualitative research designs as proposed by Roberto Hernández Sampieri (1991, p. 755). It presents features of the qualitative method because it involves "the collection, analysis and interpretation of comprehensive and [...] visual data to gain insights into" (Gay, Mills, & Airasian, 2009, p. 7) the listening assessment practices of a group of teachers from the West area of Costa Rica; also, because "the natural setting is the data source" (Tuckman, 1988, p. 388) in this study; or as Gay, Mills, & Airasian propose, in quantitative research "as much as possible, data collection occurs in a naturalistic setting" (p. 7). Our study also uses tools particular to the quantitative method to convey findings and results.

According to Hernández (1991), a mixed approach is one which combines tenets of both quantitative and qualitative approaches to conduct research. He explains that, in the past, researchers (mainly fundamentalists) believed that these two approaches were unmatchable, and that, therefore, research had to be conducted following one explicit approach. However, as he continues to explain, tendencies have changed over the past decades and, today, people see the value in using a mixed approach.

3.2 Participants

Participants in this study were fourteen MEP teachers who work in Costa Rican bilingual secondary schools of the West Area, ranked C1 according to the CEF, and who participated in a 15-week course on language assessment which included the study of listening assessment. This was a ninety-hour, teacher-fronted course offered by MEP-CONARE to MEP in-service teachers. The purpose of this training experience was to prompt the development of evaluation and assessment skills on in-service teachers so that they are able to objectively assess their students' proficiency through the English learning process. The main objective of the course was to enable participants to use theoretical and applied fundamentals of assessment to evaluate the students' acquisition and communicative use of English. Through the course both theoretical and applied principles of assessment for each language skill were studied including those pertaining to the assessment of listening skills. Thus, participants had the chance not just to discuss theory on the assessment of listening but also were able to create listening tests that were reviewed both by peers and the instructor as a way for them to become equipped with hands on knowledge on listening assessment. The course took place during the second semester of 2011.

3.3 Materials

The instrument used to analyze the teacher-created tests was a checklist (see appendix 2) previously adapted by Gamboa and Sevilla (2013a). In total, the checklist includes eleven criteria that seek to assess the degree of compliance of the tests with the theoretical principles that will be discussed in the workshop². These criteria included: test format, test heading, general test objective, general instructions, credits, balance of item difficulty, specific

² see Gamboa and Sevilla (2013b) for expansion on listening assessment principles

instructions, listening test techniques, scoring key, face validity, beneficial backwash, and listening passage.

3.4 Procedure

The first step in the design of the present research included the review of literature pertinent to the research topic in order to build a theoretical platform upon which to rest the study. Then participants were contacted to request tests that they had created and used in their actual classes (see letter of consent in appendix 1). Thus tests were collected, and then they were analyzed quantitatively by using the checklist described in subsection 3.3 above. In order to cross-check information and assure validity, triangulation was done at two different levels. First, the data resulting from the quantitative analysis were confronted with qualitative annotations in the tests. A table of codes was developed to ensure participants' confidentiality and to aid the inclusion of qualitative data. Then, the results were compared with recent studies about listening assessment in the West Area of Costa Rica. Such analyses led to the interpretation of the findings and the drawing of conclusions.

4 Analysis of the Results

4.1 Analysis of Tests created by teachers ranked C1

Data from the analysis of the tests was contrasted with the results of previous studies, as well as with current theory on listening assessment and the MEP guidelines on assessment as described in the materials provided by MEP. Results are presented in the form of graphs and tables in the pages that follow. Lastly, conclusions were drawn by contrasting results against the research question of the study.

For the sake of analysis, data were grouped into three categories according to the degree of compliance of tests with the assessment principles as dictated by the MEP and listening assessment theory. Thus, these categories are operationalized as follows: *The highest third*, which groups criteria ranking between 85 and 100%, included test heading, test format, face validity, and beneficial backwash; *the middle third*, which groups criteria ranking between 70 and 84%, comprised specific instructions and listening test techniques; and *the lowest third*, which sets criteria ranking below 69%, included general test objective, credits, general instructions, and balance of item difficulty.

4.1.1. Table of Codes

In order to ensure anonymity of the informants, the researchers have designed codes referring to the teacher created tests analyzed in the study. In the chart below, the first column depicts the informants of the study; the second one shows the name of the instrument for data collection; and the third one, the code used for the sake of anonymity in the analysis of results, where T stands for Teacher, C stands for Created, and T stands for Test. The number on the right indicates the test number. Thus, TCT-001 will refer to data gathered through the test provided by the first participant (numbers have been assigned randomly); TCT-002, to the test provided by the second one, and so forth, up to TCT-014, which will denote the test provided by participant number fourteen. The details explained herein can be summarized in table 1 below.

Table 1
Codes for Teacher Created Tests

Participant	Instrument	Codes
MEP Teachers ranked C1	Teacher Created Tests	TCT-001 / TCT-014

Source: Researchers' own design

Regarding the highest third, quantitative data show that the degree of compliance with theoretical assessment principles of test heading was 86,5%. Qualitative data, on the other hand, reveal that the areas of improvement with regard to these criteria have to do with the inclusion of data to be tested and of a line for the rater's name. As recorded in TCT-001, TCT-003, TCT-008, "areas to improve include data to be tested and a line for the rater's name". As for format, quantitative analysis depicts a total of 95, 19% degree of achievement. Qualitative data suggest that flaws in test format having to do with margins and the numbering of pages could be responsible for the tests not meeting the specifications in these criteria. As recorded in the annotations made by the researchers, "[...] margins need to be adjusted to testing requirements dictated by the MEP" (TCT-001, TCT-002, TCT-003, and TCT-008). As for the case of page numbering, TCT-010 reveals that "pages were not numbered which might cause test takers to have difficulty in following the test sequence". Lastly, 100% degree of achievement was recorded for both face validity and beneficial backwash.

Consequently, training on listening assessment positively impacted the teachers' practices regarding the categories discussed in the highest third, which suggests that teachers internalized the assessment principles to an acceptable degree. Results for this category are depicted in the figure 1 below.

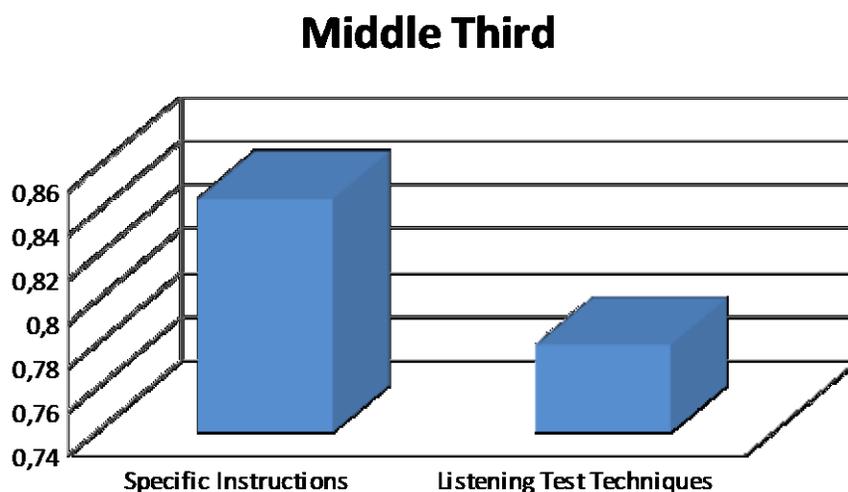


Figure 1: Highest Third: test heading, test format, face validity, and beneficial backwash

Taken together, data show generally positive results in the categories ranked as middle third. Quantitative data reveal that the degree of achievement for the compliance with assessment principles is 84,61% for specific instructions. Insights derived from qualitative analysis suggest that “not sufficient context for the task to be accomplished was provided” (TCT-002, TCT-005, and TCT-011), and also that “the total number of points and individual value of each correct item are not included” (TCT-003, TCT-004, and TCT-005), which explains the why a full degree of achievement was not reached for specific instructions. Regarding listening test techniques, the quantitative analysis depicts a total of 78% degree of achievement, while quantitative information exhibits that “advance organizers were not used to introduce each new section of the text” (TCT-001, TCT-03, TCT-005, TCT-010, and TCT-011) and that “tasks only partially reflect real-life situations” (TCT-004, TCT-005, TCT-008, and TCT-011), which, again, explains why listening test techniques have not met the desired criteria as dictated by the MEP.

The above suggests that training on listening assessment partially impacted teachers' testing practices regarding the group of criteria comprising the middle third. This implies that even though the results are seemingly acceptable, improvement needs to be made in terms of test creation, especially because both specific instructions and listening test techniques have a direct incidence on the students' performance on tests. Results for this category are depicted in figure 2 below.

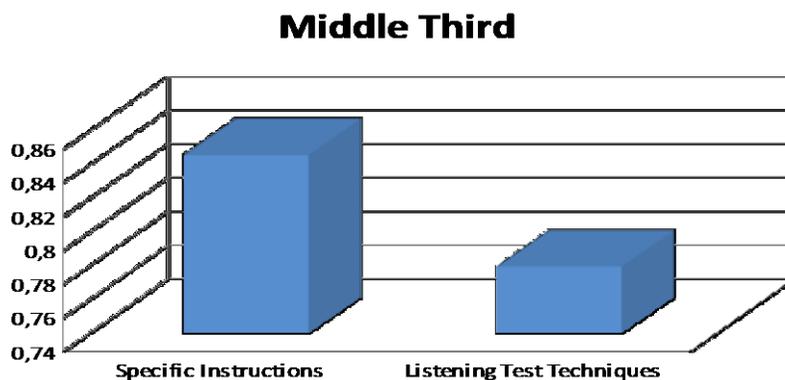


Figure 2: Middle Third: Specific Instructions and Listening Test Techniques

With regard to the lowest third, quantitative data shows that scores for the degree of achievement are 0% for general test objective; 23, 07% for credits; 30, 76% for scoring key; and 53, 85% for balance of item difficulty. By comparison, in qualitative terms the findings were as follows: Regarding test objectives, data reveals that "no test objectives were included" (TCT-012, TCT-011, TCT-008, TCT-05, TCT-04). Concerning credits, it was concluded that "[this] criterion was not observed" (TCT-001, TCT-002, TCT-004, TCT-005). With reference to scoring key, it was found that "no answer key was provided" (TCT-012, TCT-011, TCT-008, TCT-005). In relation to balance of item difficulty, it was evidenced that some of the "tests included only two parts (activities)" (TCT-003, TCT-004) which by principle hinders the achievement of balance of item difficulty.

Taken together, the data here suggest that: a) being this the group of criteria where the lowest degree of compliance was recorded, future teacher training could be oriented in this direction; and b) because balance of item difficulty directly affects student performance, improvement is paramount as a way to prompt greater chance for students' success in assessment. Results for this category are depicted in the figure 3 below.

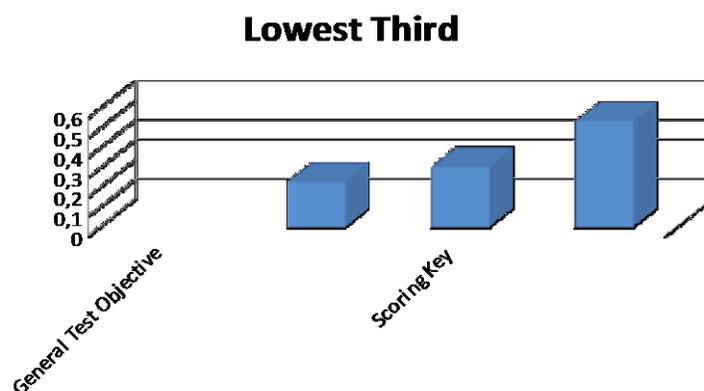


Figure 3: Lowest Third: General test objective, credits, scoring key, and balance of item difficulty

In order to provide an overall view of the degree of achievement for all criteria inquired, we have arranged the data shown in the three figures above in table 2 below.

Table 2
Overall Degree of Achievement for all Criteria Inquired

CRITERIA	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	RAW SCORE	MEAN
Test Format	87,5	87,5	87,5	100	87,5	87,5	100	87,5	87,5	87,5	100	62,5	87,5	87,5	1237,5	95,192308
Test Heading	75	75	91,6	50	75	75	66,6	91,6	83,3	91,6	91,6	100	91,6	66,6	1124,5	86,5
General Test Objectives	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
General Instructions	66,6	66,6	83,3	100	0	50%	66,6	66,6	0	0	83,3	0	83,3	83,3	700,1	53,853846
Credits	0	0	100	0	0	0	0	0	0	100	0	0	100	0	300	23,076923
Balance of Item Difficulty	0	0	0	100	0	100	0	100	100	100	100	100	100	100	900	69,230769
Specific Instructions	75	75	75	75	50	50	100	100	75	100	75	100	75	75	1100	84,615385
Listening Test Techniques	42,8	57,14	57,14	71,4	42,85	42,85	71,42	85,7	85,7	85,7	71,4	100	100	100	1014,1	78,007692
Scoring Key	100	100	0	0	0	100	0	0	0	100	0	0	0	0	400	30,769231
Face Validity	100	100	100	100	0	100	100	100	100	100	100	100	100	100	1300	100
Beneficial Backwash	100	100	100	100	0	100	100	100	100	100	100	100	100	100	1300	100

Source: Analysis of the teacher created tests.

4.2 Comparison between the results of the present study and those in the 2013 paper

This section contrasts the results in the present study and those in the 2013 paper by Gamboa and Sevilla on *The Impact of Teacher Training on the Assessment of Listening Skills*. In their 2013 paper the authors studied the correlation between teacher training and the listening assessment practices of MEP teachers of Costa Rica. To such an end, they analyzed tests created by teachers who had never received any while-in-service training on listening assessment and tests created by a group of teachers who participated in an eight-hour workshop on listening assessment.

In that paper the authors were able to conclude that tests created by teachers who had undergone training on listening assessment were significantly better than those created by teachers who received no training. They reported significant improvements in beneficial backwash, face validity, test format, test heading, and listening test techniques in the tests created by teachers who had undergone training. Arguably, results in the present study parallel the aforementioned, since, like in the former, results in the later show high compliance (between 85 and 100%) in criteria such as test heading, test format, face validity, and beneficial backwash.

Contrastingly, in their 2013 study the authors reported the need for improvement of teacher created tests in such areas as general instructions, specific instructions, general test objectives, and listening test techniques. Similarly, in the present study the authors found that general instructions, specific instructions, general test objective and listening test techniques need improvement since the ranking reported for these criteria is lower than 70%.

5 Conclusions, discussion and Implications

This study set out to examine the extent to which training on listening assessment has an impact on the actual testing practices of teachers ranked C1 in bilingual secondary schools of the West Area of Costa Rica. Based on the findings, the following conclusions are drawn:

Firstly, it is concluded that training on listening assessment has had the greatest impact on test heading, test format, face validity, and beneficial backwash. This implies that training efforts have rendered the expected outcomes in regard to these test criteria. The researchers suspect that this will translate into better assessment instruments, which, would, in turn,

provide a more accurate impression of the students' actual language competencies in secondary schools of Costa Rica.

Furthermore, training on listening assessment did not contribute to improving the areas of specific instructions and listening test techniques, general test objective, credits, scoring key, and balance of item difficulty as expected. This was contrary to expectations since, as participants had undergone extensive training that included the testing of listening, it was expected for them to rank higher in all areas. It needs to be acknowledged, though, that there were different degrees of low compliance, which ranked from 0 to 84, 61%. Thus, future teacher training should address these issues in more detail. Then, until renewed and theoretically-informed assessment practices are not incorporated, the uncertainties behind listening assessment will continue to hinder effective assessment practices in these institutions.

Another conclusion is that training on listening assessment given to date to in-service teachers has proved successful for test components such as test heading, test format, face validity, and beneficial backwash but has proven insufficient or unsuccessful for other test criteria, namely, general instructions, specific instructions, general test objective, and listening test techniques. This implies that new training efforts need to tackle the latter set of test criteria in more depth and by means of renewed strategies that prove more effective and handy to teachers as they create their listening tests. Because quality assessment can and should not respond to just some test elements at the expense of others, future training needs to equip teachers to prepare assessment instruments that comply more thoroughly with all these testing criteria.

Finally, listening assessment training programs help close the existing gap between neglected listening assessments in classrooms, MEP's lack of listening assessment guidelines, and teachers' beliefs on what listening assessment involves.

5.1 Limitations and Future Research

While this study has filled some of the gaps between theory and practice in terms of listening assessment, there are some limitations that we need to be aware of. In first place, the authors were not able to analyze the listening passage used by the participants in their tests. Also, the study did not study the correlation between test quality and teacher proficiency level. Lastly, the researchers analyzed only one test by participant, which might

not fully represent the participant's ability to create tests that meet the standards proposed by theory and by the MEP. The authors recommend that all these limitations be addressed in future research.

Future studies should also focus on three crucial areas of inquiry. The first is one deals with the assessment of other language skills such as speaking, reading, and writing. The second should look into the assessment of either listening assessment or of other skills assessment but at a national level as a way to help curricular authorities better orient their training efforts and, eventually, guide the allocation of funding for it. Lastly, the assessment of culture should be explored. This would allow the completion of the entire spectrum for language assessment within the Costa Rica's public education system context.

Over the decades subsequent to the advent of communicative approaches to language teaching and learning, listening assessment has remained the "Cinderella" of the four macro skills of English. Upon completion of this paper, the researchers propose that teacher training programs on language assessment be, in a metaphorical manner of speaking, the means to rescue such long neglected Cinderella.

6. References

- Buck, Gary (2001). *Assessing Listening*. New York: Cambridge University Press.
- Coombe, Christine A.; Folse, Keith S., and Hubley, Nancy J (2007). *A Practical Guide to Assessing English Language Learners*. Ann Arbor, Mich: University of Michigan.
- Gamboa, Roy and Sevilla, Henry (2013a). *Proceedings of the 11th Hawaii International Conference on Education: Assessment of Listening Comprehension in Public High Schools of Costa Rica: The West and Central Pacific Case*. 06-11 Jan. 2013, Honolulu, Hawaii.
- Gamboa, Roy and Sevilla, Henry (2013b). *Proceedings of the I Congreso Internacional de Lingüística Aplicada, Universidad Nacional: The Impact of Teacher Training on the Assessment of Listening Skills*. 04-06 Feb. 2013, Pérez Zeledón, Costa Rica.
- Gay, Lorraine R.; Mills, Geoffrey E. and Airasian, Peter W. (2009). *Educational Research: Competencies for Analysis and Applications*. Upper Saddle River, N.J: Merrill/Pearson.
- Hernández, Roberto; Collado, Carlos, y Baptista, Pilar (1991). *Metodología De La Investigación*. México: McGraw-Hill.

- Nunan, David. (2002). *Listening in language learning*. In J.C. Richards & W.A. Renandya (Eds.), *Methodology in language teaching: An anthology of current practice*. UK: Cambridge University Press.
- Oller, John W. (1979). *Language Tests at School: A Pragmatic Approach*. London: Longman.
- Osada, Nobuko (2004). Listening comprehension research: A brief review of the past thirty years. *Dialogue*, 3, 53-66.
- Porte, Graeme.(2010). *Appraising Research in Second Language Learning: A Practical Approach to Critical Analysis of Quantitative Research*. Philadelphia: John Benmamins.
- Presidencia de la República (2007). *Costa Rica Multilingüe. Costa Rica*.
- Tuckman, Bruce W. (1988). *Conducting Educational Research*. San Diego: Harcourt Brace Jovanovich.
- Vandergrift, Laurens. (1997). The Cinderella of communication strategies: Reception strategies in interactive listening. *The Modern Language Journal* (81), 494-505

6 Appendices

Appendix 1: Letter of Consent for Participants

University of Costa Rica
West Branch
San Ramón, Alajuela
February 01, 2013

Dear Teacher-participant:

MEP in-service English teachers ranked C1 according to the Common European Framework of Reference have been selected as participants for a research project. The study seeks to gain knowledge regarding listening assessment practices of English teachers working in bilingual middle schools from the west area of Costa Rica, who have received while-in-service training on methodology. To that end, we are requesting you to share **one listening test** that you have administered in your classes recently, along with **the listening passage** you used. The results of this study will enable us to both gain expertise as researchers in the field of listening assessment in bilingual middle schools of the west area and share insights with curricular authorities of Costa Rica in regards to future training programs for MEP English teachers. The information gathered through the analysis of your tests will be handled with absolute confidentiality and will be used for research purposes only.

This study will take place between February 1st and May 23rd, 2013. It will be conducted by M.Ed. Roy Gamboa Mena and Lic. Henry Sevilla Morales, both professors and researchers at the University of Costa Rica, West Branch. The researchers have carried out similar studies within the field of listening assessment and have shared the results in many national and international conferences. A copy of the study will be made available to you upon request once it is published on May, 2013.

By sending your listening test and the listening passage to any of the following e-mail addresses: al_deron@hotmail.com; gamboa.roy@gmail.com within the next week, you agree to participate in the present study. Please feel free to contact us if you have further questions regarding our research.

Yours, truly,

M.Ed. Roy Gamboa Mena
Professor of English
UCR- West Branch

Lic. Henry Sevilla Morales
Professor of English
UCR- West Branch

Appendix 2: Listening Test Checklist

LISTENING TEST CHECKLIST

Objective:

To assess MEP teacher-created listening tests for their compliance with theoretical principles of assessment and MEP's assessment guidelines and regulations.

Test N°: _____

LISTENING TEST CHECKLIST							
CRITERIA	TASK ACHIEVEMENT DEGREE			CRITERIA	ACHIEVEMENT DEGREE		
	YES	PARTLY	NO		YES	PARTLY	NO
Test Format				Test Heading			
1. Has the layout of the test been well set?				Are the following elements included? 1. the name of the educational institution			
2. Is it suitably and professionally arranged?				2. the school term and year			
3. Are top, bottom, left and right margins set in 2.5 cm.?				3. the type of test (midterm or final)			
4. Is the typeface style and font size big enough as to enable students to read smoothly and understand well the data included in the test?				4. data of listening to be tested			
5. Are diagrams, pictures, and other test elements well organized?				5. the total points and percentage of the test			
6. Is spacing between lines adequately set so that the test appears uncluttered?				6. the school or high school level			
7. Are all pages numbered to maintain readers well oriented on the right sequence of the test?				7. a line for the rater's name			
8. Are the photocopies clear enough for students to be able to do the exercises?				8. a line to write the date on which the test will be administered or the date is already included			

				9. the allotted time for the achievement test			
				10. spaces to set the points, grade and percentage obtained			
				11. a line for the testee to write his/her name			
				12. a line for parents to sign the test, if required			

Adapted from: MEP-CONARE– Course: Evaluación de los Aprendizajes, INSTRUMENT G

LISTENING TEST CHECKLIST			
CRITERIA	ACHIEVEMENT DEGREE		
	YES	PARTLY	NO
General Test Objective(s)			
1. Is there an evaluation objective(s) to establish what the testees should be able to demonstrate in regards to their language development?			
2. Is the objective(s) stated clearly, precisely and concisely?			
General Instructions			
1. Is the language focus on what the test takers should do rather than on what they should not do?			
2. Are instructions organized numerically or alphabetically in a proper way?			
3. Are appropriate action verbs used to introduce each set of instructions?			
4. Are explanations and/or examples specific, short and clear?			
5. Is important information highlighted whenever required?			
6. Is language adjusted appropriately to meet the students' English level?			

Adapted from: MEP-CONARE– Course: Evaluación de los Aprendizajes, INSTRUMENT G

LISTENING TEST CHECKLIST				
CRITERIA	ACHIEVEMENT DEGREE			COMMENTS Feedback based on theory and MEP's guidelines and regulations
	YES	PARTLY	NO	
<p>Credits</p> <p>Are copyright laws followed by giving credit to the authors of intellectual works such as stories, poems, illustrations, maps, and others?</p>				
<p>Balance of Item Difficulty</p> <p>Is the test arranged from the easiest to the most difficult tasks?</p>				
<p>Specific Instructions</p> <p>1. Are explanations specific, short and clear?</p>				
<p>2. Is there sufficient context for the test task to be accomplished well?</p>				
<p>3. Is the language adjusted appropriately to meet the students' English level?</p>				
<p>4. Is the total number of points and individual value of each correct item included?</p>				

Listening Test Techniques				
1. Are there appropriate test techniques to elicit those behaviours that reflect more reliably the students' specific listening abilities?				
2. Is there a minimum of three different exercises to evaluate specific listening skills?				
3. Are the questions ordered as they are heard in the passage?				
4. Are the questions spaced out in the passage?				
5. Each new section is framed with an advanced organizer to help activate the testee's schemata?				
6. Do the test tasks reflect real-life situations?				
7. Are the items placed sufficiently far apart so testees have enough time to answer one item without missing the next?				

Adapted from: MEP-CONARE– Course: Evaluación de los Aprendizajes, INSTRUMENT G

LISTENING TEST CHECKLIST	
TASK ACHIEVEMENT DEGREE	COMMENTS

CRITERIA	YES	PARTLY	NO	Feedback based on theory and MEP's guidelines and regulations
<p>Scoring Key</p> <p>Is there a scoring key specifying the acceptable answers to all listening test items?</p>				
<p>Face Validity</p> <p>Do the test content and tasks meet the objectives intended by the test designer?</p>				
<p>Beneficial backwash to</p> <p>1. Do the test content and techniques correspond with the objectives of the curriculum for which this achievement test is intended, so that its eventual administration may have a positive impact on testee?</p>				

Adapted from: MEP-CONARE– Course: Evaluación de los Aprendizajes, INSTRUMENT G