



Literatura y Lingüística

ISSN: 0716-5811

[literaturalinguistica@ucsh.cl](mailto:literaturalinguistica@ucsh.cl)

Universidad Católica Silva Henríquez

Chile

Baitman, Brittany; Véliz Campos, Mauricio

A comparison of oral evaluation ratings by native English speaker teachers and non-native English speaker teachers

Literatura y Lingüística, núm. 27, 2013, pp. 171-200

Universidad Católica Silva Henríquez

Santiago, Chile

Available in: <http://www.redalyc.org/articulo.oa?id=35228136010>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in [redalyc.org](http://redalyc.org)

[redalyc.org](http://redalyc.org)

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

# A comparison of oral evaluation ratings by native English speaker teachers and non-native English speaker teachers\*

Brittany Baitman\*

Mauricio Véliz Campos\*\*

## Abstract

This study attempts to explore the differences and similarities between native English speaker (NES) teachers and non-native English speaker (NNES) teachers in their oral evaluation ratings of the same university level English language learners. To this effect, the iBT/Next Generation TOEFL Test Independent Speaking Rubric and a questionnaire were employed. The results reveal that NES teachers are more lenient in their oral evaluation ratings than NNES teachers. In regards to the questionnaire employed, it was found that NES teachers take into consideration the aspects of fluency and pronunciation more so than NNES teachers when orally assessing students, while NNES teachers take more into consideration the aspects of grammatical accuracy and vocabulary. Further research is required in the area of oral assessment specifically pertaining to nationality, age, work experience, and knowledge of a second language.

**Keywords:** NS teacher, NNS teacher, oral assessment

# Una comparación de escalas de evaluación por profesores nativos angloparlantes y profesores no nativos angloparlantes

## Resumen

Este estudio procura explorar las diferencias y similitudes entre profesores de inglés nativos del idioma (NES del inglés) y profesores de inglés no nativos (NNES del inglés), en relación a su evaluación oral a estudiantes que cursan el mismo nivel universitario. Para ello, se utilizaron dos instrumentos: iBT/Next Generation TOEFL Test Independent Speaking Rubric y un cuestionario. Los resultados revelan que los NESs son menos severos en su evaluación oral en relación a los NNESs. En referencia a los resultados del cuestionario utilizado, éstos revelan que los profesores NES consideran más los aspectos de fluidez y pronunciación en comparación con los NNESs cuando evalúan el desempeño oral de sus estudiantes, mientras que los NNESs enfatizan más la precisión gramatical y el vocabulario. En el área de la evaluación oral aún se requiere mayor investigación, específicamente en relación a la nacionalidad, edad, experiencia laboral y conocimiento de una segunda lengua.

**Palabras clave:** profesor nativo, profesor no-nativo, evaluación oral.

Recibido: 03-03-2012

Aceptado: 27-08-2012

---

\* Profesora de Inglés, Magíster en Enseñanza de Inglés como Lengua Extranjera, profesora permanente de eClass. Santiago, Chile. [brittany.baitman@gmail.com](mailto:brittany.baitman@gmail.com)

\*\* Profesor de Inglés. Magíster en Lingüística Anglosajona. (c)Dr. *TESOL and Education*, Profesor Asociado de Universidad Católica Silva Henríquez, Santiago, Chile. [mveliz@ucsh.cl](mailto:mveliz@ucsh.cl)

## 1. Introduction

The native English speaker (NES for short) versus non-native English speaker (NNES) debate has been studied in recent years (Kachru, 1986; Canagarajah, 1999; Crystal, 2003). Frequently expressed as the NES versus NNES "dichotomy" (Liu, 1999; Medgyes, 1992), this label usually represents two mutually exclusive, opposed, or contradictory groups, yet are these two groups of speakers necessarily mutually exclusive, opposed, and/or contradictory? By referring to NES and NNES as dichotomous, greater emphasis is being placed on the differences between them, rather than the similarities that exist. Thus, it is necessary to further examine if significant differences really do exist, in this case, between NES and NNES teachers and their assessment.

The areas of the NES versus NNES debate that have been addressed most in literature are attitudes and preferences (Timmis, 2002; Watson Todd & Pojanapunya, 2009), issues of accent (Véliz, 2011), and English varieties (Kachru, 1986; Jenkins, 2006). Nevertheless, not much research has been done on NES teachers and NNES teachers and language assessment. In terms of assessment, Barrios (2002) established a distinction among these two groups of teachers, in which it was found that NNES teachers believed they had a greater capability to evaluate students' potential and to foresee their possible areas of difficulty (as cited in Madrid & Perez Canado, 2004, p. 127). However, what has not been addressed is if differences exist among these groups in terms of oral evaluation.

By examining the NES and NNES differences in relation to assessment, there is the opportunity to add to the already existing research on this debate and determine whether there are evident differences between these two groups. Research has shown that this preference for NES teachers is expressed implicitly throughout educational institutions. For instance, Watson Todd & Pojanapunya (2009) observed that English language institutions create job advertisements requiring only native speaker teachers, and institutions also boast the fact that only NES teachers are hired. This phenomenon automatically makes us and society assume that NES teachers are better in some way. The current study does not intend to label one group of teachers, NES or NNES, as better or worse. Instead, the intention is to study if there is a significant difference among the two, and this will be achieved through the evaluations of oral skills made by both groups with a holistic rubric.

Oral skills were chosen as the evaluation factor in this study for various reasons. One of which is the general interest in the skill of speaking. For instance, with any language, is it not more common to hear "Do you speak English?" rather than "Do you listen to/read/write English?" we have observed with our own students that speaking is very often the skill that wants to be practiced and that is also the most intimidating. Another reason is that there is not much current research on this specific area in terms of oral evaluation from the NES versus NNES perspective. It is not known if NES teachers and NNES teachers differ in their oral evaluation ratings. Finally, it is important to examine if NES teachers and NNES teachers use different criteria to determine a score for their students. This may give insight into which components of oral communication are more important for some teachers than for others.

This study aims to examine if NES and NNES teachers differ in their oral evaluation ratings of the same individual English language learners using the TOEFL iBT/Next Generation Independent Speaking Rubric and a questionnaire. The first section of the study deals with describing the context and participants of the study. Next, the second part of the paper defines the concepts looked at in the study and the theories surrounding those concepts. Lastly, the paper concludes with an analysis of the results and suggested areas of further research related to the concepts explored in this study.

## 2. Contextual Framework

### 2.1. Student profile

The students that participated in this study belong to a private university consisting of about eight thousand three hundred students in total. The university is located in the district of Peñalolén, in a remote area at the foot of the Andes Mountains. The facilities are modern and the resources abundant. In 2011, this university was ranked the number one business school in all of Latin America by Ranking América Latina.

The English program for the students in this study consists of levels one (lowest) to seven (highest). The students are placed into a level following a diagnostic test. This test follows the guidelines established by the institution, according to which "evalúan en forma progresiva tu nivel de Inglés de acuerdo a las habilidades de: comprensión lectora, au-

ditiva y conocimiento gramatical<sup>1</sup>. "The students are given one chance to complete the test and sixty minutes to do so. The diagnostic test is one hundred percent multiple choice consisting of listening, reading comprehension and grammar questions. Therefore, the test does not take into account speaking or writing. Students must pass their current level to continue to the next level. To pass levels one to six, they must have over a 4,0 average (Chilean scale) on a combination of two quizzes, two oral evaluations, online work, class participation, and one final exam. In order to pass level seven, the students must take the TOEIC® and receive a score of 700 or above. If the students do not pass the TOEIC® with a score of at least 700, they must continue with English courses in level eight, which is designed as a TOEIC® preparation course and re-take (and pass) the TOEIC® at the end of the course. Every student at the university must pass the TOEIC® in order to graduate.

## 2.2. Rater profile: Native English speaker teachers

The six NES teachers participating in this study were chosen because of their gender, educational backgrounds and nationality. All these English teachers possess an accredited *Teaching English as a Foreign Language* certificate. Their work experience includes teaching in institutes, businesses, schools, and universities in Santiago, Chile. They were selected due to their nationality, in order to have a more diverse sample, and gender, in order to have an equal amount of both men and women. All the NES teacher raters are from the United States, United Kingdom, or Australia.

## 2.3. Rater profile: Non-native English speaker teachers

As the NES teachers are determined "native" based on the first language they learned, the NNES teachers in this study are considered "non-native" on the basis of their L1 not being that of English. The six NNES teachers participating in this study were chosen because of their similar educational backgrounds and are all Chilean. All these teachers hold university degrees in the area of Education from Chilean universities and are working towards a Masters in TEFL or have already obtained a Masters degree in TEFL or a similar area. Their work experience is very similar to the NES teachers and includes that of institutes, businesses, schools,

---

1 The source remains anonymous in order to safeguard the privacy of the institution.

and community colleges in Santiago, Chile. They were also selected because of their gender, in order to have an equal amount of both men and women.

### 3. Theoretical framework

#### 3.1. Introduction

In order to understand why a study is dedicated to examining differences between NES and NNES teachers and assessment, one must first be aware of the theoretical foundations of these topics. There are a number of reasons why such a distinction has become a TESOL issue, namely the spread of English and its geopolitical implications, the importance of speaking as a skill, and the use of oral evaluations. The following section includes a review of the pertinent literature with regards to the two main themes of this study: the NES versus NNES debate and assessment. Also, the findings of current studies touching on these themes are also discussed. The following section begins with implications as to why there is a distinction between NES and NNES teachers in the first place.

#### 3.2. The spread of English

The beginning of the spread of English can be dated all the way back to the fifth century, when English began its stretch around the British Isles after its arrival in England from Northern Europe (Crystal, 2003). Throughout the hundreds of years following, many significant events occurred, particularly the founding of the "New World" in the seventeenth century, which began the launch of English to the global status it holds today. According to Crystal (2003):

The present-day world status of English is primarily the result of two factors: the expansion of British colonial power, which peaked towards the end of the nineteenth century, and the emergence of the United States as the leading economic power of the twentieth century.(59)

The United States has a major influence in how English is developed throughout *the* world due to having the largest number of NESs and the economic power it holds (Crystal, 2003). In order for other countries to do business with the dominant economic power, the inhabitants must learn the language of that country, in this case, English.

English has become the language to learn in the world today, and it is wholly evident. In fact, the amount of NNEs throughout the world has already outnumbered the amount of NESs. Also, the spread of English has created a world full of different varieties of English that are primarily used in the outer and expanding circles. For instance, Kachru (1986) claims that the English in the outer circle has become institutionalized or "nativized" due to its use in the educational and legal systems, and therefore the English in those areas has transformed, creating new norms that are not only used, but accepted (McKay, 2002). However, "standard English" has not yet disappeared and is used by those mainly in the inner circle countries. According to Quirk (1990), standard English "is what might be termed the unmarked variety; it is not unusual or different in any way and is typically associated with written English" (as cited in McKay, 2002, p. 51).

Since these varieties are used by more speakers than "standard English," why do the varieties not hold as much value as does standard English? It is apparent that English varieties have gained at least some regard due to the coining of the term "world Englishes" and the influence of the work done by Kachru, also known as the Kachruvian approach. Through this approach, Kachru claims that the concept of "world Englishes" "emphasizes 'WE-ness,' and not the dichotomy between us and them (the native and non-native users) (Kachru, 1992; as cited in Bolton, 2006).

### 3.3. NES teacher versus NNE teacher debate

Davies (1996) asserts that "The native speaker is a fine myth: we need it as a model, a goal, almost an inspiration. But it is useless as a measure; it will not help us define our goals" (as cited in Medgyes, 2001, p. 157). This idea leads more towards the modern way of viewing the NES teacher versus NNE teacher debate, according to which, we need some type of standard that motivates us to reach our highest potential, but that our ability cannot be measured by that standard alone. If we analyze this in terms of English language teaching (ELT), NNEs should aim for a native-like competency, but their competency does not determine how good of an English teacher they are.

When revisiting Quirk (1990) and standard English, he also asserts that "the implications for foreign language teaching are clear: the need for native speaker support and the need for non-native teachers to be in

constant touch with the native language" (as cited in Bolton, 2006, p.251). Here, we see some support for NNES teachers, but only on behalf of native-speaker support and native-speaker English as the model. Quirk basically claims that NNES teachers are just fine, *as long as* there is a NES at their side. Conversely, there has not been much literature dedicated to what the NNES teacher can bring to the table that NES teachers cannot. As a way of illustration, it is not rare that a NES teacher is not familiar with the L1 of his or her students. If we are constantly concerned with the needs of our students and the context in which we teach, wouldn't the native language of our students play an immense role in our teaching? Widdowson (2003) shares his opinion on the matter:

If we acknowledge that TESOL - the teaching of English to speakers of other languages - cannot effectively be done without reference to the other languages, that guiding the development of bilinguals has to be attuned to the bilingualization process, and not by the imposition of an exclusive monolingual pedagogy, then there is some hope at least that English can be learnt without denying the legitimate rights of less privileged minority languages. (162)

Widdowson makes another point by mentioning that it is not only knowing the L1 that is helpful, but being aware of the "bilingualization" process. In other words, in order to teach another language, the teacher should be familiar with how a second language is acquired. What better way to have done this than for the English teachers to have personal experience and have learned another language themselves? Clearly the NNES teachers have that advantage as they know and speak their L1 and an L2 (English). However, this cannot be said nor is it required of all NES teachers.

### 3.4. Speaking as a skill

The process of speaking includes three main phases: conceptualizing the message content, formulating the message linguistically, and articulating the message (Bygate, 1987). It is supposed that these three phases work together effortlessly and create clear, coherent speech. Nevertheless, problems can arise at each of the three phases. For instance, when conceptualizing the message content, the speaker can give an inappropriate message. In the second phase, when formulating the message lin-



guistically, the speaker can choose the wrong word. Finally, when articulating the message, the speaker can pronounce some words incorrectly. These problems can happen to speakers in their L1, so for L2 speakers, these problems become significantly more likely.

Speaking is an essential part of learning a language. Three main theories exist as to what is needed for speakers to speak in a foreign language and for which aspects of speaking are most important. The three models are the Speaking as a Process model (Bygate, 1987), the Communicative Language Ability model (Bachman & Palmer, 1996; as cited in Alderson & Bachman, 2004), and Activity Theory (Vygotsky). The first two models mentioned above will now be discussed in detail, while Activity Theory will be discussed further in the following section entitled *Assessing Speaking in L2*. In his model of speaking as a process, Bygate (1987) makes the distinction between knowledge and skill. Consider the following claim:

We do not merely know how to assemble sentences in the abstract: we have to produce them and adapt to the circumstances. This means making decisions rapidly, implementing them smoothly, and adjusting our conversation as unexpected problems appear in our path. (Bygate, 1987, 3)

It is clear that for Bygate, it is not just what a learner knows about the L2 that will help him or her to produce speech. He affirms that *knowledge* is what allows learners to speak, and *skill* is the component involved in actively engaged interaction. "By giving learners 'speaking practice' and 'oral exams' we recognize that there is a difference between knowledge about a language, and skill in using it" (Bygate, 1987, p. 6). According to Bygate, one is not more important than the other; both knowledge and skill are needed when any person speaks.

The communicative language ability model by Bachman and Palmer (1996; as cited in Alderson & Bachman, 2004) is one of the most regularly used communicative models in language testing today (Alderson & Bachman, 2004), and is made up of five components; the first of which is called language knowledge. Language knowledge is the most complex, as it consists of a range of types of knowledge that the speaker may have and can be broken down even further into two more categories: organizational and pragmatic. Organizational knowledge clearly concentrates on organization and consists of grammatical knowledge and

textual knowledge. The first type, grammatical knowledge, refers to vocabulary, syntax, phonology, and graphology, whereas the second type, textual knowledge, refers to cohesion and conversational organization (Alderson & Bachman, 2004).

### 3.5. Assessment

In terms of assessment, two of the most common forms are summative and formative, diagnostic being the least discussed. Summative assessment is defined as formally planned assessments which are usually used to evaluate student progress and/or grade students and is usually given at the end of a unit, semester, or year (Davison & Leung, 2009). Formative assessment differs from summative assessment in that it is more informal, frequent, and carried out while students are learning. These two types of assessment have generally been viewed as very different. However, researchers such as Kennedy et al. (2006; as cited in Davison & Leung, 2009, 398) have pointed out that summative assessments can also be used for formative purposes, and that the division between these two assessment types is no longer necessary. This is shown in their inclusive model for assessment, in which Kennedy et al. (2006; as cited in Davison & Leung, 2009, 398) propose the following:

1. All assessment needs to be conceptualized as assessment for learning.
2. Feedback needs to be seen as a key function for all forms of assessment.
3. Teachers need to be seen as playing an important role not only in relation to formative assessment but in all forms of summative assessment as well—both internal and external.
4. Decisions about assessment need to be viewed in a social context because in the end they need to be acceptable to the community.

As seen in number three, Kennedy et al. (ibid.) claim that teachers play an important role in all forms of assessment. Simply put, the way in which a teacher maneuvers assessments in the classroom and the teacher's perceptions or beliefs can affect the outcome of the assessment and the students themselves. For instance, let us take the case of a holistic oral assessment rubric. If the teacher feels that one aspect of language is more

important than another, the final score will be weighted differently, and that student will receive a different score than he or she would have with another teacher. If the way in which English teachers rate their students in an oral evaluation have significant affects on students, it is important to determine if differences in teacher raters exist and where exactly these differences lie among these teachers.

### 3.5.1. Assessing speaking in L2

Assessing speaking can be difficult in ELT, due to the fact that a teacher's impression on how well a learner speaks can be based on many different factors. It is argued that students should be assessed in a "real" situation in which "real" language must be used. According to activity theory, stemming from the ideas of Lev Vygotsky, mental behaviour is seen as action (Lantolf, 2000, as cited in Alderson & Bachman 2004). The focus here is not on the individual learner, but on the activity. In activity theory, "activities are considered significant when the individual acts purposefully in order to accomplish some goal" (Alderson & Bachman, 2004, p. 102). Many speaking assessments are structured in this way. They are meant to be meaningful for the learner and also require "real" language so that some final goal is reached. One limitation involved with speaking assessments is actually having to simulate this real-language use situation, which can lead to many complexities, namely context and costs. Activity theory emphasizes the fact that oral assessments are also activities, and the participants (both teachers and students) will interpret it as an activity. In other words, they will know it is an assessment, and this will influence how the assessment is carried out by both teachers and students.

As mentioned previously, the communicative language ability model (Bachman & Palmer, 1996; as cited in Alderson & Bachman, 2004) has been found useful in creating a framework for speaking assessments (Alderson & Bachman, 2004). Although language knowledge is just one component of the model of language ability, it is often given the most weight in speaking assessments used by English teachers worldwide.

As has been shown, the production of speech includes a lot more than just knowing a language, and there are a series of processes involved. By examining the processes of speech, it is more easily understood why oral assessments must contain different criteria when assessing a student's

oral skills. There is a lot involved, so that must shine through on the scale being used to assess.

### 3.5.2. Oral assessment instruments

The instruments used to assess speaking provide scores that "express *how well* the examinees can speak the language being tested" (Alderson & Bachman, 2004, 59). Commonly expressed as numbers, the scores are usually associated with a set of descriptors in order to have a better idea of what the number stands for. The numbers and sets of descriptors ranging from high to low are what make up a rating scale for assessing speaking (Alderson & Bachman, 2004). Just as there are two common forms of assessment in general, there are two common types of speaking scales as well. Holistic scales consist of the rater's impression based on a set of descriptors and one overall score, while analytic scales contain multiple criteria assigned to numbers with a set of descriptors for each.

### 3.6. Current studies in oral assessment by NES and NNES teachers

In a study by Zhang & Elder (2011), an oral evaluation (the College English Test-Spoken English Test [CET-SET] of China) was administered to thirty test-takers in China. A total of ten speech samples were collected, each containing three speakers. The data collected for the study came from the holistic numerical ratings given by nineteen NES and twenty NNES teacher raters as well as comments made by the raters regarding the scores they gave on the CET-SET for each participant. The results showed no significant differences between the holistic ratings of the NES and NNES teachers. However, although the holistic ratings did not differ much, the comments made by the raters showed differences in which criteria used to come to those final ratings held more importance. Linguistic features were found to be more important for the NNES teachers than for NES teachers, while the NES teacher comments were more widely distributed among other categories of oral communication such as demeanour, interaction, and compensation strategy (Zhang & Elder, 2011). This shows that the NES teachers considered more criteria than the NNES teachers when assessing the students. Consider the following conclusion:

The NES raters' intensive focus on the Interaction and Compensation Strategy categories, for example, indicated that the NES raters tended to make a judgment on how well the candi-

dates can meet the requirements of a communicative task that they might be required to carry out in a real world encounter, while the NNES raters focused more on test-takers' underlying language ability as manifested through task performance. (Zhang & Elder, 2001, 43)

These findings are especially relevant, as the current study will also examine the holistic ratings made by the teachers and which features of oral communication are more important for NES and NNES teachers. Although some of the objectives are very similar to the present study, the results from Zhang & Elder's study are most pertinent to the Chinese context. The NNES participants all come from Chinese universities, and the oral evaluation scale used to assess the learners (CET-SET of China) is specific to Chinese universities. In studies such as the one presented above, it is still unclear whether raters place more or less importance to students' speech depending on the level of students being assessed.

We will be able to compare the results from the current study in a Chilean context to the results from a similar study in a Chinese context. In the next section, the methodology of the current study is discussed in detail and is then followed by the results.

## 4. Methodological framework

### 4.1. Research methodology and methods

This study incorporates elements of positivism in that it declares objectivity; it aims to explain, not understand. In this study, the reality being researched is "captured" and is undisturbed by the methods. In other words, the researcher is looking at this reality from the outside. For that reason, positivism is present in this study.

The instrument being used to evaluate students' speaking skills is the Independent Speaking Scale used on the internet-based (iBT)/Next Generation TOEFL Test, which is holistic rating scale. The TOEFL test measures both receptive and productive skills (listening, reading, speaking, and writing). On the speaking section, the examinee has twenty minutes to complete six tasks. The first two tasks are "independent" tasks, which focus on topics considered familiar to the examinee. The other four tasks are "integrated," as the test taker needs to use more than one skill in order to produce the answer. For instance, in one of the tasks, the examinee

has to listen to a short talk and question, and then answer the respective question with an oral response. In this task, the examinee needs to use his or her listening skills in conjunction with speaking skills. The two independent tasks use the independent rating scale, which differs from the integrated rating scale used for the four integrated tasks. The examinee uses a microphone and headset to complete the speaking section. When completed, the responses are digitally recorded and sent off to ETS's On-line Scoring Network. The digitally recorded responses are then evaluated holistically by three to six certified raters on a scale of zero to four. Those scores are then converted to a zero to thirty scale (ETS, 2005b).

In this study, only the independent rating scale is used by the raters. After signing the consent form, each of the three students is given four independent tasks to complete. A blank consent form and an example of a completed consent form have been annexed. Two of the tasks ask the student to express a personal choice and defend that choice regarding certain topics they are familiar with. The other two tasks ask the student to agree or disagree with an idea, then express and defend that choice. For each task, the students are given one minute to prepare and a maximum of two minutes to respond and are able to take notes during the preparation time. The students' responses are recorded onto a recording device and played for the raters at the time of evaluation. The tasks for the students to complete have been annexed and are the following:

1. Describe a place you would like to visit and explain why you chose that location. Use details and examples to support your decision. (Response time: 30 sec. - 45 sec.)
2. Describe a class you have taken in school and explain why the class was important to you. Use details and examples to support your decision. (Response time: 45 sec. - 1 min.)
3. Agree or disagree with the following statement: University education in Chile should be free. Use details and examples to support your decision. (Response time: 1 min. 30 sec. - 2 min.)
4. Agree or disagree with the following statement: All university students should pass an English language exam (e.g. TOEIC) in order to graduate. Use details and examples to support your decision. (Response time: 1 min. 30 sec. - 2 min.)

After signing the consent form, the six NES raters and six NNES raters are given a copy of the iBT/Next Generation TOEFL Test Independent Speaking Rubric and are asked to study the descriptors in the scale. The iBT/Next Generation TOEFL Independent Speaking Rubric has been annexed. They are informed of the study and of how they will be evaluating three English language learners using the given rubric. At the time of evaluation, the raters are given a copy of the Independent Speaking Rubric and a scale of zero to four for each student. The scale has been annexed. The students being evaluated are assigned letters and are referred to as Student A, Student B, and Student C. The teachers must circle a score of zero (speaker makes no attempt to respond OR response is unrelated to the topic) to four (the response fulfils the demands of the task with at most minor lapses in completeness) for each student according to the criteria found on the Independent Speaking Rubric.

After evaluating Students A, B, and C with the Independent Speaking Rubric, the raters must complete a questionnaire regarding the ratings they gave in their evaluations. The questionnaire has been annexed. The questionnaire uses a Likert scale of importance and is designed to measure which aspects of oral communication influence the ratings made by the NES and NNES teachers. The scale takes into account the following linguistic components of oral communication: accuracy, fluency, pronunciation and vocabulary. Definitions for each of the components are provided and are read by the teachers before completing the Likert scale of importance. The raters must consider how important each component of oral communication was for them when giving their scores on the Independent Speaking Rubric. Then, they circle a score of one (Very Unimportant) to five (Very Important) for each component.

## 4.2. Sampling procedure

As mentioned in the Contextual Framework, the teachers were chosen because of certain criteria they possess (NES/NNES status, gender, nationality, and experience with advanced ELLs). Therefore, the purposive sampling method is used. Purposive sampling involves selecting participants in order to elicit the data in which the researcher is interested (Mackey & Gass, 2005). Nationality was used as a criterion to choose the teachers in order to have NESs from various countries, and experience

with advanced learners was a criterion due to the fact that the speech samples rated by the teachers are of advanced ELLs. The teachers selected were emailed with detailed information on how the study would be conducted and what is required of them as participants. When a teacher confirmed his or her participation, a time and place were coordinated to meet and begin the tasks.

The sample of teachers was made up as follows: six NS teachers, three males and three females, and six NNS teachers, three males and three females.

As far as choosing the students who provide the speech to be assessed by the teachers, they were chosen using the convenience sampling method. First, levels 7 and 8 were chosen as the levels to select the students from, primarily because they seem more capable of engaging in oral interaction. Then, all the students from that level were given the opportunity to participate. The first three students to respond were selected. In contrast to the purposive sampling method used to choose the teachers, the convenience sampling method chooses participants based on who is available to participate (Mackey & Gass, 2005). Levels 7 and 8 were chosen, because these are the highest levels in this particular English program, and at these levels, the students are expected to hold a conversation in various situations and orally defend their points of view using fluent, coherent speech. The students, as well as the teachers, signed a consent form, and their participation was entirely voluntary. The students' grades were not affected by participating in this study and no type of reward was given to the students or the teachers.

#### 4.3. Research questions

- Do NES and NNEST teachers differ in their oral evaluation ratings of the same university level English language learners using the iBT/Next Generation TOEFL Test Independent Speaking Rubric and a questionnaire?
- Are NES teachers more lenient in their oral evaluation ratings than NNEST teachers?
- Are certain components of oral communication more important for NES teachers than for NNEST teachers when orally assessing ELLs?



### 4.3.1. General objective

To explore the differences and similarities between NES and NNES teachers in their oral evaluation ratings of the same university level English language learners using the iBT/Next Generation TOEFL Test Independent Speaking Rubric and a questionnaire.

### 4.3.2. Specific objectives

- To analyze the oral evaluation ratings between NES and NNES teachers.
- To investigate if certain components of oral communication are more important for one group of teachers (NESs or NNESs) than the other.

## 5. Results & Discussion

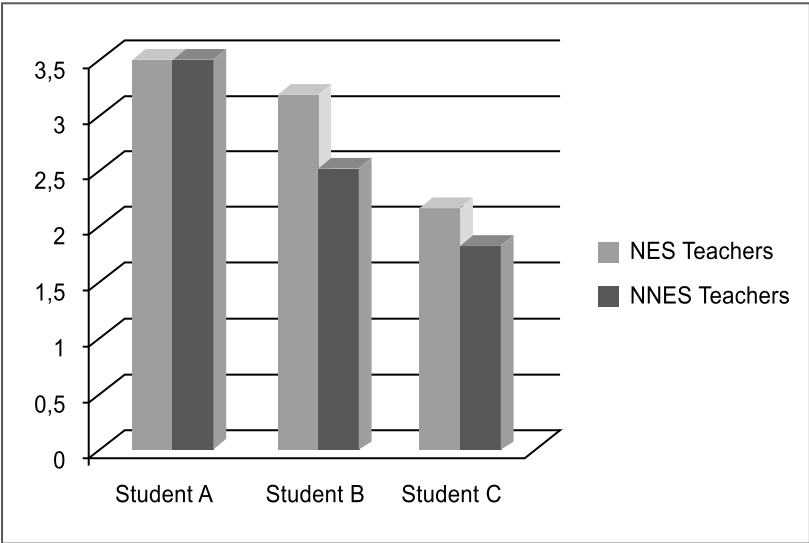
### 5.1. Oral assessment results

The data gathered in this study came from two groups of teachers: NESs and NNESs. The two instruments used to gather the data were the iBT/Next Generation TOEFL Independent Speaking Rubric and a questionnaire. On the speaking rubric, the teachers had to give a score of zero (speaker makes no attempt to respond OR response is unrelated to the topic) to four (the response fulfils the demands of the task with at most minor lapses in completeness). The results from the speaking rubric are given below and found in the figures that follow.

The ratings given by the NES teachers and NNES teachers for Student A are exactly the same. As a group, the mean score given was a 3.5. The standard deviation (SD) is also exactly the same (0.5). For Student B, on the other hand, a higher mean score of 3.17 was given by the NES teachers, whereas the NNES teachers gave a mean score of 2.5. The SD of the NES scores was higher at 0.69, and for the NNES was again 0.5. As for Student C, a higher score was again given by the NES teachers. The NES teachers gave an average score of 2.17 with a SD of 0.69, while the NNES teachers had an average score of 1.83 with a SD of 0.69 as well.

	Student A		Student B		Student C	
	Mean score	SD	Mean score	SD	Mean score	SD
NES teachers	3.5	0.5	3.17	0.69	2.17	0.69
NNES teachers	3.5	0.5	2.5	0.5	1.83	0.69

**Figure 5.1.1. Oral assessment ratings by NES & NNES teachers**



**Figure 5.1.2.Oral assessment ratings by NES & NNES teachers**

The results of the oral assessment ratings show that there is a difference in how NES and NNES teachers orally assess ELLs. Although both groups of teachers had the same average score for one student (Student A), the average ratings for the other two students (B & C) were different. For Students B & C, the NNES teachers gave lower scores than the NES teachers. The SDs did not differ much between groups. However, the more reliable results were those of Student A by both groups and of Student B by the NNES teachers, due to the fact that the SDs were only 0.5, whereas the SDs for the rest of the results were higher at 0.69. Therefore, both the NES and NNES teachers agreed most on the score for Student A.

In Figure 5.1.3., the ratings are displayed by gender (men and women) for both the NES and NNES groups. Although gender was not included as a specific objective in the research question, the findings in regards to gender are worth sharing. For Student A, the men gave an average score of 3.33 with a SD of .47, while the women gave an average score of 3.67 with the same SD of .47. The same mean score of 2.83 was given by both men and women for Student B; however the SD for men was .90 and for women was .37. As for Student C, the men and women both had the same mean score of 2, yet the SD varied (men .58, women .82).

	Student A		Student B		Student C	
	Mean Score	SD	Mean Score	SD	Mean Score	SD
Men	3.33	.47	2.83	.90	2	.58
Women	3.67	.47	2.83	.37	2	.82

Figure 5.1.3. Oral assessment ratings by male and female teachers

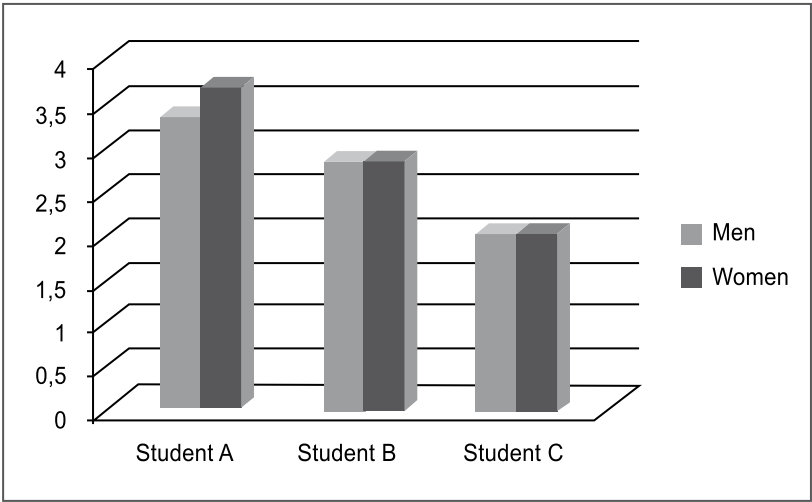


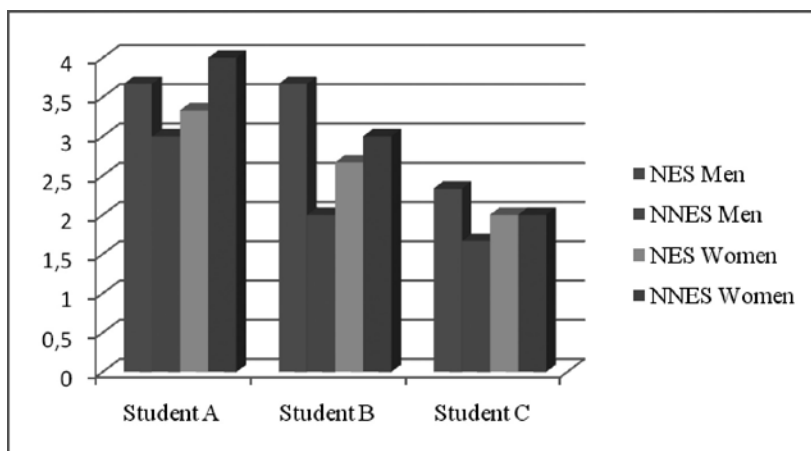
Figure 5.1.4. Oral assessment ratings by male and female teachers

The average ratings given by men and women, regardless of their NES/NNES status, are quite similar. Student B and Student C received the same average score, 2.83 and 2.0, respectively, by the men and women. Student A on the other hand received a higher average rating of 3.67

by the women, in comparison to 3.33 by the men. Therefore, there were no significant differences between the male/female groups when NES/NNES status was not taken into account. Again, the SDs for Student A were exactly the same (and quite low) for both the men and women, demonstrating that the teachers in each group were in high agreement on the scores for Student A. However, with Student B, the SD for the men is staggering (0.90), in comparison to the SD for the women (0.37). The men's SD shows a lot of dispersion in the scores, which makes the mean score of 2.83 very misleading. This is interesting due to the fact that the mean score for Student B was exactly the same for the men and women, yet when reviewing the SDs, it is clear that the mean score for the women is a lot more reliable than the mean score for the men. The opposite occurred with Student C, although the difference was not as great as with Student B. The mean score was exactly the same between the men and women, yet the SD was higher this time for the women (0.82) than for the men (0.58). In this case, the mean score of 2 is more reliable for the men, as there is less dispersion among the individual scores. These results are of course taking into account only gender. When NES/NNES status was also taken into account with regards to gender, disparities were more apparent. The mean scores and SDs are listed in Figure 5.1.5. below.

	Student A		Student B		Student C	
	Mean Score	SD	Mean Score	SD	Mean Score	SD
NES men	3.67	.47	3.67	.47	2.33	.47
NNES men	3	0	2	0	1.67	.47
NES women	3.33	.47	2.67	.47	2	.82
NNES women	4	0	3	0	2	.82

**Figure 5.1.5. Oral assessment ratings according to gender and NES/NNES status**



**Figure 5.1.6. Oral assessment ratings according to gender and NES/ NNES status**

When examining the speaking rubric ratings in terms of gender and NES/NNES status, the two groups that gave the highest average scores in general for all students were the male NESs and the female NNESs. The lowest scores for each of the three students were given by the male NNESs. When comparing the differences in average scores between the NES and NNES men and the NES and NNES women, the differences were greater and a lot more significant between the NES and NNES men. For instance, Student B is the most extreme case. The highest average score was given by the NES men (3.67), while the lowest average score was given by the NNES men (2). Also noteworthy from this set of results are the SDs of the NNES males and females, specifically for Students A and B (SD of 0). Since the SD is 0, all the teachers in each group gave the same exact score. There was no dispersion whatsoever. This means the average scores given by these two groups for Students A and B are very reliable. In fact, the NNES men, who gave the lowest mean scores for each student, also had the lowest SDs. Thus, the lowest average scores that were given by the NNES males are actually the most reliable of all the groups of teachers. So, when averaged on the basis of only gender, the speaking rubric results are almost identical. Yet, when examining the gender groups in terms of NES/NNES status, the results are quite varied.

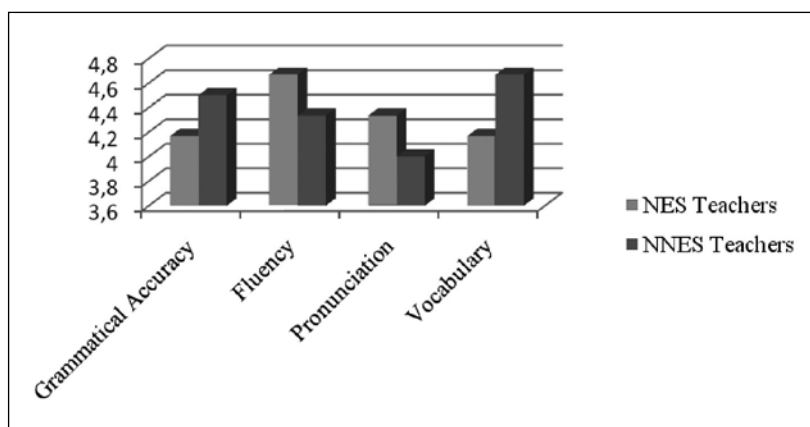
### 5.2. Questionnaire results

All of the teachers completed a questionnaire regarding different components of oral communication that they consider important for them when orally assessing ELLs. This questionnaire uses a Likert scale of importance. The scale ranges from 1 (Very Unimportant) to 5 (Very Important). The results from the questionnaire are given below and found in Figures 5.2.1. and 5.2.2.

The average score given by the NES teachers in the category of grammatical accuracy was 4.17, with a SD of 0.69, whereas the NNEs gave a higher average score of 4.5, with a SD of 0.5. On fluency, the NESs gave a mean score of 4.67, with a SD of 0.47, and the mean score given by the NNEs was a 4.33, with the same SD of 0.47. The average score given by the NESs for pronunciation was 4.33 with a SD of 0.47, while the average score given by the NNEs was 4 with a SD of 0.82. For vocabulary, the NESs gave a mean score of 4.17 with a SD of 0.37, whereas the mean score given by the NNEs was a 4.67 with a SD of 0.47.

	Grammatical accuracy		Fluency		Pronunciation		Vocabulary	
	Mean Score	SD	Mean Score	SD	Mean Score	SD		
NES teachers	4.17	0.69	4.67	0.47	4.33	0.47	4.17	0.37
NNEs teachers	4.5	0.5	4.33	0.47	4	0.82	4.67	0.47

**Figure 5.2.1. Importance of components of oral assessment by NES & NNEs teachers**



**Figure 5.2.2. Importance of components of oral assessment by NES & NNES teachers**

As is seen in Figure 5.2.2., NES teachers scored fluency and pronunciation as more important when assessing students. For NNES teachers, grammatical accuracy and vocabulary were more important. It is interesting to note that the NNES teachers scored the two components that are more related to language knowledge (grammatical accuracy & vocabulary) as more important than the NES teachers. The NES teachers, on the other hand, scored the two components that are more related to language use (fluency & pronunciation) as more important than the NNES teachers.

### 5.3. Discussion

As the NES teachers generally rated the students higher on the speaking rubric than the NNES teachers, these results coincide with the common presumption that NES teachers are more lenient than NNES teachers. If we consider the Zhang & Elder (2011) study mentioned earlier in Section 3.7, similar results were not found. The nineteen NES and twenty NNES teachers gave holistic scores of 1 (Very poor) to 5 (Excellent) of thirty participants' speech samples. In the Zhang & Elder study, no significant differences were found between the NES and NNES teacher ratings. This could of course be due to the context, as the Zhang & Elder study was carried out in China, in a different institution, and with a different style rubric. As the current study compared NES teachers with only Chilean NNES teachers, the results may only be applicable to the Chilean context.

With regards to gender and scores, there were no significant differences between the male/female groups when NES/NNES status was not taken into account. When NES/NNES status was taken into account, the average scores between men and women were more wide-ranging. This is quite interesting, because, for instance, when the average scores for Student B were looked at only in terms of the gender of the raters, the scores were exactly the same between the women and men. This makes one think that they basically agreed on the scores, and there is no difference between the men and women for this student. Nevertheless, after filtering the groups based on NES/NNES status, the two groups of men have very different scores. It can be concluded from these findings that NES/NNES status has a greater impact on scores than does gender.

The results from the questionnaire are noteworthy in terms of oral assessment. It is clear that the NES and NNES teachers hold preferences for different aspects of oral communication when assessing ELLs. These results are in some ways similar to the results found in the study by Zhang & Elder (2011), in which holistic numerical ratings supplemented by comments were given by nineteen NES and twenty NNES teacher raters for each participant. Although the results showed no significant differences between the holistic ratings of the NES and NNES teachers, the comments made by the raters showed differences in which criteria used to come to those final ratings held more importance. Overall, linguistic features were found to be more important for the NNES teachers than for a NES teacher, which seems to be similar to the results of the current study. Yet when examining how the researchers defined "linguistic resources," the results change slightly. The "linguistic resources" category is divided into four more categories: phonology, vocabulary, grammar, and general linguistic resources. The only significant differences in comments between NES and NNES teachers are in the vocabulary and general linguistic resources categories. Even though the NNES teachers made more comments in regards to the linguistic resources category, the NES teachers mentioned vocabulary significantly more often than the NNES teachers. This differs from the current study in that the NNES teachers rated vocabulary as more important than the NES teachers. Also in the Zhang & Elder study, the NES teachers mentioned fluency more than the NNES teachers. This is similar to the findings in the current study, but this finding was not found significant in the Zhang & Elder study. Many conclusions can be drawn not only from the results of the current study, but from the entire investigation, and are discussed in the following section.



## 6. Conclusions & projections

This research study has hopefully shed light on the myths and realities of the NES/NNES teacher debate, and on the fact that there are differences in the assessment styles of NES and NNES teachers. What is important to understand is that this is okay. One style is not necessarily better than the other. Just as students have different learning styles, teachers have different assessment styles, and some styles work better for some students and situations than for others. This study came about from the common assumption that NES teachers are less demanding than NNES teachers in terms of assessment. This research study aimed to determine if that really is the case by examining the oral evaluation ratings and questionnaire results of the same individual English language learners using the TOEFL iBT/Next Generation Independent Speaking Rubric and a questionnaire by both NES and NNES teachers.

As mentioned in the previous section, the results show that the NES teacher participants in this study gave higher scores than the NNES teacher participants. This seems to support the assumption that NES teachers are more lenient than NNES teachers when orally assessing ELLs. However, the results from this study differ from a very similar study by Zhang & Elder (2011) mentioned in various sections of the paper. In Zhang & Elder's study, the ratings made by the NES and NNES teachers were not significantly different, which demonstrates that there may be more at stake than just NES/NNES status. Consequently, the ratings were compared on the basis of gender. There were no significant differences in the average scores of the males and females. As a result, the groups were then compared on the basis of both the prior categories: gender and NES/NNES status. When these two categories were both taken into account when comparing the ratings, the average scores were different, and there was a distinct difference between NES men and NNES men. Although the conclusion that NES teachers are more lenient than NNES teachers in oral assessment can be drawn from the results from the current study, these results have been contradicted in other studies from other contexts. Therefore, the possibility exists that other aspects may have a role in determining how one orally assesses ELLs, and this would be an excellent area to research further.

There are many more factors included in this study that could have influenced the results obtained. One aspect that could have influenced

the teachers' ratings is the educational background of the teachers. In this study, all of the NNES teachers possess a university degree in the teaching of English. This contrasts from the NES teachers, who also possess a university degree, but not in the teaching of English. Instead, the NES teachers only have a Teaching English as a Foreign Language (TEFL) certificate. The educational backgrounds, then, are very different among the teachers. Another factor that could have played a role in the results is not only the educational background, but the language background of the NES teachers. It has been found that English teachers that have learned a second language themselves are more familiar with the acquisition of a second language. A teacher with experience in learning a second language, and even more so in learning the students' mother tongue, may assess differently than would a teacher with no personal experience in learning a second language. The fact that the NESs focus more on fluency and pronunciation definitely tell us something: they are more focused on communication. This is usually associated with a NES orientation. Oftentimes, when a NES teacher is living and teaching in a foreign country and is obligated to speak the language of that country just to get by from day to day, communicating what he or she needs is the most important. Another possibility for the focus on communication is that not all NESs arrive in the foreign country already knowing the language. In fact, many arrive with no prior exposure to the language at all. Therefore, they eventually "pick-up" on the language and acquire it, rather than learning through grammar drills. This could be a factor for why the NESs put less emphasis on grammar and vocabulary and more on the communication aspect. Medgyes (1994) affirms, NES teachers "do not make a fuss about errors unless they hinder communication" (as cited in Lasagabaster & Sierra, 2002, 136). The focus on communication has been supported in recent years. However, this emphasis on communication by NES teachers has actually been found detrimental to students (Lasagabaster & Sierra, 2002).

Furthermore, it has been found that language teachers with knowledge of a second language prefer to teach in the same way that they themselves were taught. In this case, the communication aspect can be well-explained. The same goes for the NNES teachers, which in this study *had all* learned English first in the school, so they have that knowledge base. For that reason, they could have used that knowledge when orally assessing the students. The fact the NNES teachers possess this linguistic

knowledge base of English contrasts very much with the NESs in that the NES teachers did not learn English in the same way. Just as many of them have acquired the language of the country in which they teach, they also, as in the case of almost any L1, acquired English. They may actually be less familiar with grammatical structures than NNES teachers. All these factors mentioned would be useful to consider in further research.

Another area stemming from the current study that would be interesting to research is the differences in the ratings of NNES teachers on the basis of nationality. For instance, do Chilean, Chinese, and Indian NNES teachers differ in their oral evaluation ratings of ELLs? Even age would be an attractive area, which was not taken into consideration in this study. All the teacher participants were under the age of thirty-five. The results, therefore, may be distinct from those of older teachers, and possibly with more experience. Also on the topic of experience, the teacher participants do not work in the same institutions, and therefore work with different curriculums. Some curriculums may have a greater emphasis on the receptive skills, for instance, than the productive skills. If this is the case, some teachers may have more or less experience than others in terms of speaking skills and this could play a role in the results. Thus, it would be practical to compare the speaking rubric results of NES and NNES teachers with significantly more experience in the field of speaking with the results of NES and NNES teachers with less experience. Although one must use the rubric as a guide while assessing, it would be interesting to see if the results show a difference depending on experience. These are just a few more areas of research that can stem off the current study leading to a greater pool of knowledge in the topic of L2 oral assessment.

The most important conclusions to be drawn from this work are the implications for teachers and students of English, and for educational institutions in Chile. It is clear that students and many institutions still hold the presumption that the NES model is best. In other words, they prefer NES teachers to NNES teachers. Many scholars have worked to break this myth and many studies have been done to shed light on the benefits of NNES teachers and also the disadvantages of NES teachers. This study does not deem one "better" than the other, yet it does show differences between the two when it comes to assessment. How one interprets the differences is up to that person in particular. It is possible for a particular institution, for instance, to prefer English teachers to be more severe when assessing their students. In this case, according to the results of this

study, it may be best to hire NNES teachers. On the other hand, maybe an institution's English department is already made up of both NES and NNES teachers. In this case, it may be best to form a collaboration type of oral assessment, including a combination of NES and NNES teachers. According to the questionnaire results of this study, NES and NNES teachers consider different components of oral assessment more important than others. If these two groups of teachers are grading students with more consideration for one aspect than would another teacher, students will most likely receive different scores depending not on their actual skills, but on the teacher. If a NES teacher that values fluency and pronunciation orally assesses a particular student along with a NNES teacher that values grammatical accuracy and vocabulary, a more reliable assessment would be made than if the assessment came from just one of the teachers alone. In fact, some institutions around the world have now started implementing a team-teaching technique, in which a NES teacher and NNES teacher share a classroom (Lasagabaster & Sierra, 2002). This can be found in parts of Spain and Japan, among other countries, yet has not been implemented into the Chilean educational system yet.

The NES/NNES debate continues to be a hot-topic in the world of TESOL. As this world is constantly changing, we as teachers and researchers need to keep up with the transformations and accept rather than deny them. This means adapting our methods and approaches to take into account the ever-changing areas of TESOL, such as the growing number of NNESs and NNES teachers and English varieties. It is a thing of the past to assume that the NES model is best and discriminate teachers on this behalf. We must be able to recognize that there are advantages and disadvantages to both NES and NNES teachers. Both should be respected and valued just the same in TESOL. There will always be "good" and "bad" teachers, but should never be associated solely with their NES/NNES status. In the end, this will only help us to be better professionals and better teachers and will undoubtedly lead to more success in the classroom.

## 7. References

- Alderson, J. C., & L. F. Bachman (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.
- Bolton, K. (2006). World Englishes today. In B. B. Kachru, Y. Kachru, & C. L. Nelson, *The handbook of world Englishes* 240-269. Oxford: Blackwell.
- Bruthiaux, P. (2010). World Englishes and the classroom: An EFL perspective. *TESOL Quarterly* 44(2): 365-369.
- Bygate, M. (1987). *Speaking*. Oxford: Oxford University Press.
- Canagarajah, A. S. (1999). Interrogating the "native speaker fallacy": Non-linguistic roots, non pedagogical results. In G. Braine, *Non-native educators in English language teaching*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.: 77-92.
- Celce-Murcia, M., D. M. Brinton & J. M. Goodwin (1996). *Teaching pronunciation*. New York: Cambridge University Press.
- Cook, V. (1999). Going beyond the native speaker in language teaching. *TESOL Quarterly* 33(2):185-209.
- Crystal, D. (2003). *English as a global language* (Second ed.). Cambridge: Cambridge University Press.
- Davidson, K. (2007). The nature and significance of English as a global language. *English Today* 23 (1): 48-50.
- Davison, C., & C. Leung (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly* 43 (3): 393-415.
- ETS. (2005a). *TOEFL iBT scores*. Princeton, NJ: Educational Testing Service.
- \_\_\_\_\_. (2005b). *TOEFL iBT tips*. Princeton, NJ: Educational Testing Service.
- He, D., & Q. Zhang (2010). Native speaker norms and China English: From the perspective of learners and teachers in China. *TESOL Quarterly* 44 (4): 769-789.

- Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford, England: Oxford University Press.
- \_\_\_\_\_. (2006). Current perspectives on teaching world Englishes and English as a lingua franca. *TESOL Quarterly* 40(1): 157-181.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk, & H. Widdowson, *English in the world*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. (1986). *The alchemy of English: the spread, functions, and models of non-native Englishes*. Oxford, England: Pergamon Press.
- Kramsch, C., & W. S. Lam (1999). Textual identities: The importance of being non-native. In G. Braine, *Non-native educators in English language teaching*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.: 57-72.
- Lasagabaster, D., & J. M. Sierra (2002). University students' perceptions of native and non-native speaker teachers of English. *Language Awareness* 11(2): 132-142.
- Liu, J. (1999). From their own perspectives: The impact of non-native ESL professionals on their students. In G. Braine (Ed.), *Non-native educators in English language teaching*. Mahwah: Lawrence Erlbaum Associates, Inc.: 159-176.
- Mackey, A., & Gass, S. M. (2005). *Second Language Research*. Mahwah: Lawrence Erlbaum Associates, Inc.
- Madrid, D., & Perez Canado, M. L. (2004). Teacher and student preferences of native and non native foreign language teachers. *Porta Linguarum* 2: 125-138.
- McKay, S. L. (2002). *Teaching English as an international language*. Oxford: Oxford University Press.
- Medgyes, P. (1992). Native or non-native: Who's worth more? *ELT Journal* 46(4): 340-349.

- Medgyes, P. (2001). When the teacher is a non-native speaker. In M. Celce-Murcia, *Teaching English as a second or foreign language*. Boston: Heinle & Heinle: 429-442.
- Seidlhofer, B. (Ed.) (2003). *Controversies in applied linguistics*. Oxford: Oxford University Press.
- Timmis, I. (2002). Native-speaker norms and international English: a classroom view. *ELT Journal* 156 (3): 240-249.
- Véliz Campos, M. (2009). The nature of educational inquiry. Exeter: Unpublished doctoral assignment.
- Véliz Campos, M. (2011). A critical interrogation of the prevailing teaching model(s) of English pronunciation at teacher-training college level: A Chilean evidence-based study. *Literatura y Lingüística* 23: 213-236.
- Watson Todd, R., & Pojanapunya, P. (2009). Implicit attitudes towards native and non-native speaker teachers. *System* 37:23-33.
- Widdowson, H. G. (2003). *Defining issues in English language teaching*. Oxford: Oxford University Press.
- Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English speaking teacher raters: Competing or complementary constructs? *Language Testing* 28 (1): 31-50.