2016

An Investigation into Native and Non Native English Speaking Instructors' Assessment of University ESL Student's Oral Presentation

Rubaiyat Jabeen
Minnesota State University Mankato

An Investigation into Native and Non Native English Speaking Instructors' Assessment of

University ESL Student's Oral Presentation

By Rubaiyat Jabeen

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Master of Arts

in

English

Teaching English as a Second Language

Minnesota State University, Mankato

Mankato, Minnesota

July 2016

Date: <u>July 11th, 2016</u>

This thesis paper has been examined and approved.

Examining Committee:

_____

Dr. Glen T. Poupore, Chairperson

_____

Dr. Sarah A. Henderson Lee, Committee Member

Abstract

This thesis aims to conduct a mix-methods investigation into native English-speaking teachers (NESTs) and non-native English-speaking teachers (NNESTs) assessment of university level English as a second language (ESL) student's oral presentation. To collect data for this study, all faculty members and instructors currently teaching in at the departments of English and Communication Studies at Minnesota State University, Mankato (MNSU) were invited to participate in an online survey using Qualtrics Survey Software (Qualtrics). After receiving email invitations, altogether 31 people provided their consent to participate. Among the 31 teacher-participants, there were 19 NESTs and 12 NNESTs. The participants then took part in the data collection process by completing the online survey where they firstly watched and assessed a video-recorded oral presentation of an ESL student by scoring her performance using an analytic rating scale and then provided feedback on the student's strengths and areas of weakness. Secondly, each participant rated a question-item section on the survey about which oral presentation assessment criteria they thought was most or least important to them as raters. Finally, the participants answered five biographic questions to provide required personal information. The data was analyzed using Qualtrics and SPSS 23 Software. To begin the data analysis, the participants were divided into two groups, i.e. group 1 NEST and group 2 NNEST on the basis of the participant's biographic data provided on Qualtrics. Afterward, SPSS was used to calculate descriptive statistics such as the mean (M) and standard deviation (SD) of the data, and to conduct a t-test to compare the data from the two groups and identify any significant difference between them. Next, to determine the inter-rater reliability, the Intraclass Correlation Coefficient on SPSS was used to estimate the Cronbach Alpha for: 1), all the participants; 2), the NEST participants; and 3) the NNEST participants. As it is a mixed-methods research study, the

qualitative data, in the form of the feedback provided by the participants, were categorized according to which assessment criteria the participants commented on the most and the least, how many participants from the two groups commented on them, and how many times they were acknowledged to be strengths or pointed out as areas of improvement. The result of the qualitative data analysis was then compared to the results of the quantitative data analysis. In case of the quantitative data, the results revealed that even though there were differences in NESTs and NNESTs' assessment of the oral presentation, the differences did not hold statistically significant value. For example, judging by the mean scores on the assessment criterion 'speaks naturally', there seemed to be a noticeable difference between the NEST and NNEST groups. However, when an independent samples t-test was performed, it resulted in no significant difference. The reliability statistics disclosed that the Cronbach's Alpha of the scores given by all participants (0.72) and the NEST participants (0.71) represented good inter-rater reliability. However, the Cronbach's Alpha of the scores given by the NNEST participants (0.19) showed low inter-rater reliability. The determination of which assessment criteria were regarded as most or least important in the perception of the participants resulted in a significant difference between the two groups in their rating of the importance of oral presentation as an effective speaking assessment tool, where the NNESTs agreed that it was important but the NESTs less so. In the end, the analysis of the qualitative data revealed areas of consistency and also inconsistency in the feedback in comparison to the quantitative data. For example, the analysis of the qualitative data revealed that a high number of NESTs commented on the assessment criteria regarding natural speech and most of the comments were positive. On the other hand, most NESTs provided negative comments on 'speech volume'. This result is inconsistent with the quantitative result as the highest mean score in NEST group was received by natural speech and

the lowest was received by speech volume. Likewise, in the NNEST group, a high number of participants commented on 'pronunciation' and' body language' and most of those comments were negative. A low number of NNEST participants commented on 'eye contact' and that comment was negative. This result is partly consistent with the quantitative data because the lowest mean score in the NNEST group was given to 'pronunciation' on the rating scale. One of the limitations of this study is the low number of participants. It is suggested that a larger group of NESTs and NNESTs should be invited in case of future research. In addition, the study also puts forth pedagogical implications. For instance, it discusses the need to train raters prior to the assessment process, as well as the need for NNEST raters to know how to provide constructive feedback on grammar.

Table of Contents

List of Tables

Dedication

To:

My loving Tanzil Malek, my beautiful daughter Tanisa Diyanah Malek and my beloved

grandfather Justice Mahfuzur Rahman

Acknowledgements

This dissertation would not have been possible without the assistance of my family, mentors, colleagues and friends. The credit for this degree is shared, but the faults in it are all mine.

First and foremost, I would like to express my sincere and heartfelt gratitude to my academic advisor Dr. Glen Poupore and my second reader Dr. Sarah Henderson Lee for helping me to see this project through. I appreciate the long hours of consultations and feedback that you have offered. Any balance that this study shows is reflective of the diverse perspectives offered by the committee members and their commitment to a quality product.

I am also immensely thankful to Dr. Karen Lybeck and Dr. Nancy Drescher for their incessant support, continuous encouragement, wise guidance, constructive feedback, and endless patience throughout my MA-TESL studies.

I would also like to take this opportunity to give many thanks to my wonderful, thoughtful and enduring husband, Tanzil Malek. Your encouragement, love, and support have gotten me through my degree, and for that I could not thank you enough. You have done more than your share of housework and babysitting to give me the time to complete this goal.

In addition, I thank Minnesota State University, Mankato (Department of English - TESL Program) for the teaching assistantship opportunity I was given. The assistantship was an invaluable experience both professionally and personally. I thank the students I was privileged to teach and from whom I learned much as well.

I could not thank my best friends Cezara Talmaci and Nigina Boltaeva enough. They have truly been the support system I leaned on all through my degree. I am also thankful to my

classmates and friends Kadidja, Latifah, Ngongose, Yuka, Hiroko, Dorie, Firdavs, Shyam, Ryo and Charlotte for their friendship and cooperation.

Finally, I would like to convey my appreciation to all the faculty members and instructors of the departments of English and Communication Studies who willingly completed the online survey for this study.

Many thanks to everyone who, in one way or another, helped me to complete this research study. May God bless you all!

Chapter I

Introduction

Background Information

The English language is the most widely acquired second language all across the globe today. For more than two decades, it has been an unchallenged lingua franca, and the most common means of global communication (Crystal, 1997; Kachru, 1992; Medgyes, 1994; Prodromou, 1992). According to a TESOL International Association Annual Report (2014), there are 1.5 billion English learners worldwide. The number of English language learners (ELL) is rapidly growing in the contexts of both English as a foreign language (EFL) and English as a second language (ESL). In addition, according to the Open Doors Report on International Educational Exchange (2014), the total number of international students in the colleges of the United States has grown by 72% since 2000. International students now make up about 4% of all university students in the United States of America. For example, according to the Kearney International Center at Minnesota State University, Mankato (MNSU), there are more than 900 international students from 90 different nations at this university. Each semester there are approximately 100 students who take ESL courses (Minnesota, 2015a). As a part of this development, the interactive and oral communicative purposes of English at the US colleges and universities are being emphasized increasingly. All four skills of the English language, namely, speaking, listening, reading and writing are now developed with equal importance in building communicative competence. Standardized English language proficiency tests such as the Test of English as a Foreign Language (TOEFL) and the International English Language Testing System (IELTS) now have a speaking component that is crucial in determining a learner's level of

language competency and knowledge. Similarly, in college level English language classrooms, the speaking and listening components are integral as well.

Owing to this reputation that is given to communicative competence, and by extent to oral communication, there has been a shift from traditional pedagogies to the adoption of Communicative Language Teaching (CLT) in the field of English language teaching and learning. A learner is said to have communicative competence when he/she is able to interpret and perform appropriate social activities in his/her target language (Canale and Swain, 1980; Celce-Murcia et al., 1995; Hymes, 1972). Since the term "communicative competence" was coined by Hymes (1972), CLT has attracted many researchers and language teachers. CLT is an approach to language teaching that is based on the theory that the primary function of language is communication (Brandl, 2008).

Consequently, as more and more language teachers chose to implement CLT practices in their classrooms, by placing the development of communicative skills at the forefront and introducing grammar only in relevance to the development of the skills, the organization of the course syllabi began to undergo new changes. Instructors began incorporating task-based instruction (TBI) into their syllabi, with which tasks became the principal units constituting the basis of daily and long-term lesson plans (Brandl, 2008). In addition, keeping in line with alternative approaches and methods to language teaching, the assessment practices also evolved, and alternative assessment started becoming popular among language teachers.

Huerta-Macías (1995) refers to alternative assessment as a substitute for standardized testing. Alterative assessment is considered to nurture productive learning instead of being ends in themselves. A popular example of alternative assessment is performance-based assessment. According to McBrien & Brandt (1997), performance-based assessment is when the teacher is

measuring how well a student can do on authentic and real world-like tasks that require them to communicate. Apart from emphasizing the assessment of language in communicative forms, performance-based assessment allows the teachers to assess the process as well as the product of students' learning.

One of the most widely applied types of performance-based assessment is oral presentation. Joughin (2010) defines oral presentation assessment as an evaluation of learning on the basis of the student's spoken skills. In oral presentations, the mode of communication between student and teacher is the various forms of spoken words. The ESL courses at MNSU such as, 1) ESL 125: Advanced Oracy Skills for Non-Native Speakers, and 2) ESL 102/103: Intensive Oral Skills for Non-Native Speakers, make great use of oral presentations to assess their ELLs. However, oral presentation as an assessment tool has been criticized by many researchers for its lack of reliability. Rater reliability, more specifically inter-rater reliability, is crucial in the assessment of speaking. Inter-rater reliability refers to the consistency in scores of the raters. Unclear scoring criteria, fatigue, carelessness or a bias toward 'good' and 'bad' students can all play a part in a rater's own reliability (Brown, 2007). Joughin (2010) states that as raters generally have to score the presenters as the presentation proceeds, it could introduce bias. Brown (2007) mentions that for an assessment process to be high on reliability it needs to be consistent and dependable. The topic of inter-rater reliability is closely linked to the scoring practices of the raters/teachers. It has to be mentioned here that due to the given trend of globalization, not only is the number of international ELLs but the number of international English language teachers (ELTs) is also rapidly increasing. Because of this rapid increase, the inconsistency of assessment practices and inter-rater reliability of native English speaking

teachers (NESTs) and non-native English speaking teachers (NNESTs) is a very crucial area of research.

According to the Teaching English as a Second Language (TESL) list of faculty and staff, currently, there are 16 ELTs at MNSU (Minnesota, 2015b). Largely, ELTs can be divided into two groups. On the one hand, there are native English speaking teachers (NESTs) and on the other hand, there are non-native English speaking teachers (NNESTs). The major distinguishing factor between the NESTs and the NNESTs is that for the former, English is their first language (L1) and for the latter, English is not their L1. Out of the 16 ELTs at MNSU there are eight who are non-native speakers (NNS) of English, and the other eight are native speakers (NS) of English. The diversity in the educational and professional experiences of the teachers makes their practices and method of scoring dissimilar too. Studies indicate that raters' scores are affected to some degree by rater attributes such as age, gender, occupation, international experience, personality, cultural background and opinion (Barnwell, 1989; Eckes, 2008; Galloway, 1980; Kang, 2008; Ludwig, 1982; Reed & Cohen, 2001). Research studies also suggest that the group of NNESTs is becoming more populous than their counterparts (Braine, 2005; Samimy & Brutt-Griffler, 1999). This growth in the population of NNESTs in the field of English language teaching has brought about an increasing focus on the teachers themselves (Llurda, 2005).

However, the limitation of those studies is that their topics are repetitive (Braine, 2005). The majority of the studies investigate the differences of teaching practices between NESTs and NNESTs. Among the limited number of studies that were carried out on the differences of assessment practices, most shed light on the assessment of writing, leaving the area of speaking assessment unexplored. That creates a great scope and need for research focusing on the differences in the assessment practices of speaking between the NESTs and NNESTs. Kim

(2009b) states that the qualitative analysis of her study demonstrated that the judgments of the NESTs were more detailed and elaborate than those of NNESTs in regards to pronunciation and grammar. Nevertheless, while some researchers found that differences exist between NESTs and NNESTs, others argue that there is no strong evidence to support the differences (Brooks, 2013). Therefore, this study examines if NESTs and NNESTs rate the same oral presentation differently or not, and also focuses on the inter-rater reliability among the participants.

Purpose and Rationale

The purpose of this study is to investigate the differences in the assessment practices between NESTs and NNESTs in relation to an oral presentation of a university ESL student in the US. In the opinion of O'Sullivan (2012), "tests of spoken language ability is the most difficult to develop and administer" (p. 234). While there is a collection of studies that investigate the various aspects of assessing speaking, such as discussing how the characteristics of the raters, namely personality, age, gender and experience, impact assessment (Berry, 2004; O'Sullivan, 2000, 2002, 2008), there are not too many studies that take the language background of the raters into consideration. With the increase in the number of international teachers who are NNS of English, it is essential to investigate whether there are different speaking assessment practices between NESTs and NNESTs, and if yes, what they are. Moreover, it is particularly essential to compare the assessment practices of NESTs and NNESTs in the ESL context. Since the majority of NNESTs teach in their countries of origin or in the EFL context, limited studies have looked into NNESTs teaching in the ESL environment where English is the first language (McNeil, 2005). However, as the number of NNESTs in the ESL context is gradually increasing

(Liu, 2005), it is necessary to analyze the assessment practices of NNESTs and examine how they compare to those of the NESTs in the ESL teaching environment.

As international students applying to the United States for higher education, my husband and I were obliged to take the IELTS together. During the speaking component of the test, I was interviewed by a NNS of English while my husband was interviewed by a NS of English. Even after having studied in English immersion schools throughout my life, and as a result being more proficient in communicating in English compared to my husband, the spoken test scores of my husband were higher than mine. This incident led me to wonder if NNS speaking test examiners exercised more strictness in their scoring practices than their NS counterparts. It also made me speculate about the inter-rater reliability of the test takers, which I assumed was low.

After coming to MNSU, and being afforded the opportunity to work as a Graduate Teaching Assistant (GTA) in the ESL context, I have taught speaking skills to ELLs alongside other NESTs. As an ELT who is prudent about the self-improvement of personal teaching skills, one of my major questions and area of interest has always been how I as a NNEST compared to my fellow NEST colleagues as far as assessment practices where concerned. Moreover, I also thought it was crucial to find out the extent of inter-rater reliability among all the instructors who were teaching the same speaking courses to different groups of ELLs at MNSU.

My interest to investigate the difference in the speaking assessment practices between NESTs and NNESTs was further intensified when I began to take notice of how my husband and his NS equivalents varied in their scoring of their students' oral presentations. My husband is a GTA in the department of Communication Studies at MNSU. The prime responsibility of the Communication Studies GTAs, most of whom are NS of English, is to teach Speech Communication courses to NS American students as well as NNS international students. Having

had the opportunity to observe the Communication Studies GTAs closely, I noticed how most NNS GTAs are inclined to score the grammar and form of the students' speech while the NS GTAs prefer to focus on body language and pronunciation. After this intriguing observation I decided to conduct my research to explore how and why the NESTs and NNESTs are different in their assessment practices of oral presentations. What is more, the decision to conduct this research became conclusive on reviewing related literature, which revealed a need for more studies in this area.

Research Questions

The purpose of this study is to investigate how NESTs and NNESTs assess ESL students' oral presentations. To that end, the specific research questions are as follows:

1. Are there any differences between native English speaking (NS) and non-native English speaking (NNS) instructors in the way they assess English as a second language (ESL) students' oral presentations in a US university context? If so, what are those differences?

2. What is the inter-rater reliability among:

    i. All the participants;

    ii. Among NS participants, and

    iii. Among NNS participants?

3. Which assessment criteria does each group think are more important or less important?

Structure of the Thesis

This thesis is comprised of five chapters. The next chapter reviews some existing literature related to the topic of this study. The chapter firstly starts with a discussion of the comparison between NESTs and NNESTs, including their definitions, language proficiency,

features of language teaching, and NEST and NNEST as raters of oral performance. Secondly, the chapter looks at issues in the assessment of speaking, taking into account test takers or presenters, task performance parameters, rating scales and rater performance. Thirdly, the chapter reviews literature on the importance of oral presentations as an assessment tool for ESL students.

Chapter three focuses on the methodology on this study. It describes the research setting and participants, research materials, data collection process, and data analysis of the quantitative and qualitative data.

Chapter four presents the analysis of the results while Chapter five discusses the main findings of the study. In addition, Chapter five concludes with the limitations of the study, suggestions for future research, and the researcher's observations and pedagogical implications.

Chapter II

Literature Review

This chapter reviews some of the existing literature on topics related to the assessment of speaking and the differences between NESTs and NNESTs. More specifically, the topics that have been explored here are firstly, a comparison between NESTs and NNESTs, including their definitions, language proficiency, features of language teaching, and NESTs and NNESTs as raters of oral performance. Secondly, the chapter looks at issues in the assessment of speaking, taking into account test takers or presenters, task performance parameters, rating scales and rater performance. Thirdly, the chapter reviews literature on the importance of oral presentations as an assessment tool for ESL students. The review of these studies has helped in identifying the gap that the current study should attempt to fill. It has informed the current study about the existing scenarios in the field of speaking assessment pertaining to NESTs and NNESTs, including the challenges and limitations of conducting a research in this area.

Comparison between NESTs and NNESTs

From the time of the creation of the Non-native English Speakers' Caucus in the TESOL International Association in 1999, studies on NNESTs have been conducted abundantly. Those studies focused particularly on the areas of language expertise, teaching performance and teaching approaches (Braine, 2005; Canagarajah, 1999). Some studies have also been conducted to find out about the NNESTs' personal perceptions about their differences from NESTs, and their preferences and the challenges they meet (Braine, 2005; Boyle, 1997; Liu, 1999; Medgyes, 1992; Prodromou, 1992). Later studies regarding NESTs and NNESTs have also focused on

their different assessment practices (Brooks, 2013), but they do not effectively explore the reasons behind the differences.

Definition Criteria of NS and NNS of English

Commonly, in the field of English language teaching, the people who speak and use English as their L1 are categorized to be NS of English, while the people for whom (regardless of their competence) English is not their L1 are known as NNS of English. Thomas Paikeday in his book, *The Native Speaker is Dead!* (1985) recorded a discussion with more than forty prominent linguists about their perceptions of the NS individuals. The derived definition of NS from his book stems from the method of language acquisition that states, "a person who has a specified language as the mother tongue or first learned language" (Paikeday, 1985, pp. 9-10). However, not all scholarly definitions are as clear-cut. According to Chomsky (1965), it is unlikely for NS individuals to forget their L1, while NNS individuals have an unsteady state of their target language resulting in loss or forgetfulness of the language after learning it. However, Paikeday (1985) criticizes that explanation with the argument that it is a very metaphysical and ambiguous definition of English speakers' minds.

In McKay's (1992) opinion, NS and NNS should be defined on the basis of the speakers' countries of origin. The countries or regions of origin can be categorized into three circles, namely inner, outer and expanding (Kachru, 1992). Firstly, the inner circle countries are the ones where people use the English language as their L1 daily and for all purposes. Secondly, the outer circle countries use English as their second language (L2) or for official purposes. Thirdly, the expanding circle countries are where English is learned as a foreign language but not used officially. Therefore, the people from the inner circle countries are to be defined as NS and the

rest as NNS (McKay, 1992). Nevertheless, McKay and Karchu's suggestions are not free of criticisms. Some scholars (Liu, 1999; Medgyes, 1994) contend that the English language has now been established as a universal language and is used for communication globally. Additionally, according to Liu (1999), many people from outer circle countries are bilingual and have native-like proficiency, which makes it unreasonable to define them as NNS.

Even after reviewing the conflicting and contradicting definitions of NS and NNS, the dichotomy is being used in this study because it is still the most familiar and convenient terminology to distinguish between the two groups. The term NS is used in order to identify the speakers for whom English is their L1 and NNS for the speakers for whom English has not been their L1. Correspondingly, NESTs are NS English teachers, while the NNESTs are NNS English teachers.

Language Proficiency

Can NNS individuals attain a native speaker-like proficiency level in Second Language Acquisition (SLA)? Studies in SLA often compare and contrast the performances of the two groups in order to be able to answer this question (Brooks, 2013). According to Brooks (2013), no fundamental difference exists between NS and NNS users of English. Nonetheless, there is debate regarding this issue. As per the theories of SLA, differences in the language acquisition process can have an impact on the language user's proficiency (Finegan, 2004). Owing to the different learning processes, the language adaptability of the NNS users may be less than NS users, who on the contrary are believed to have superior perception and control of the language (Freeman & Freeman, 2003). Furthermore, the NS users are generally expected to have a better knowledge of the complexities of language such as lexicon and idiomatic expressions as opposed

to the NNS users (Medgyes, 1994). Respectively, a number of studies have been carried out to inspect how different NESTs and NNESTs are in terms of language proficiency. As stated by Baratt & Kontra (2000), NESTs have more language authenticity, for example their pronunciation is accurate, their range of vocabulary is wide and they have full cultural understanding of the language. Comparatively, NNESTs may be less competent speakers, with language that sounds bookish (Medgyes, 1994).

Most of the differences between NESTs and NNESTs that were found through further research relates to language authenticity. According to Medgyes (1994), NNESTs are not likely to use English as resourcefully as NESTs. For many NNESTs, especially the ones working in the EFL setting, opportunities and methods to use English in real-world like interactive situations are not easily available. Therefore, learning English as a foreign language may impede elements of pronunciation such as stress, intonation and rhythm, and overall fluency of speech. According to Llurda (2005), NNESTs are more cautious when giving lectures because of their accented English, which makes them difficult to follow at times.

In contrast, despite the less authenticity in language use and less native-like speaking skills, the different language learning processes makes NNESTs more adept in certain linguistic knowledge compared to NESTs (Finegan, 2004). As explained by Finegan (2004), "second-language learners ordinarily have linguistic meta-knowledge that is lacking at least in the early stages of a first language" (p. 520), because the L1 learners acquire the language naturally and subconsciously. Therefore, NNESTs have a more conscious understanding of how the elements of lexicon, phonology, grammatical forms and discourse of their target language are different from their L1.

More often than not, the experience of learning L2 renders the NNESTs to be more insightful about grammatical accuracy and functions than NESTs (Celce-Murcia & Larsen-Freeman, 1999). For that reason, NNESTs are motivated to exert more importance and interest in English grammar both as users and teachers of the language.

In his study that surveyed approximately 100 supervisors of NNS Teaching English as a Second Language (TESL) graduate students in the United States to explore their language proficiency as future NNESTs, Llurda (2005) found evidence of NNESTs having better linguistic knowledge. Based on the supervisors' regular observation and assessment of the NNS TESL graduate students, the survey reported that they had equal or better linguistic knowledge than their NS counterparts. More precisely, 50% of the survey participants stated that the TESL students as future NNESTs had equal ability, and 34% stated that the TESL students as future NNESTs had higher ability in comprehending and explaining the English language system at all teaching levels.

To summarize, it can be said that both NESTs and NNESTs have certain strengths and drawbacks in terms of language proficiency. Although the NNESTs are more prone to using textbook-like unauthentic language, with limited competency in pronunciation, they have the advantage of being knowledgeable about and proficient at teaching the grammatical system of the language.

Different Features of Language Teaching

It is not improbable for the abovementioned differences in language proficiency between NESTs and NNESTs to lead to differences in their language teaching practices. According to Canagarajah (1999), the rather complex process of language teaching has many aspects, starting

from communicative competence, linguistic knowledge, teaching experience, professional training to personal preparation. In Braine's (2005) opinion, the majority of the different teaching practices employed by NESTs and NNESTs correlate to their differing language proficiency levels. Therefore, the teaching focus, use of language and teacher-student relationship for NESTs and NNESTs will vary in accordance with their distinguished characteristics as language learners and users of English themselves.

Studies (Lasagabaster & Sierra, 2002; McNeil, 2005) demonstrate that NESTs and NNESTs have different teaching focuses. For example, NNESTs are more inclined to focus on and give feedback to students' language errors. Correcting students' language errors is more natural with NNESTs because of their better awareness of the language system and personal learning process. In a survey conducted by Medgyes (2001), 82% of the NEST and NNEST participants confirmed that there were dissimilarities existing in their teaching practices. Through data collection, Medgyes found that unlike NESTs, NNESTs maintained strictness regarding language error correction and provided extensive amounts of corrective feedback. In the same way, Lasagabaster & Sierra (2002) hold that NESTs have the tendency to exercise greater leniency towards student's language errors as long as communication is effective.

NESTs are also more likely to adopt CLT approaches, while for many NNESTs, CLT proves to be challenging due to their unauthentic language usage and limited competency in speaking (McNeil, 2005). As already mentioned, NESTs' ease and expertise with the colloquial form of the English language enables them to work interactively with the students, and to facilitate them to acculturate with the authentic and idiomatic forms of the language more confidently. In contrast, Medgyes's (2001) study found out that instead of CLT, NNESTs use

approaches that focus on the language mechanism by incorporating controlled activities on forms, grammar rules, registers and accuracy.

However, a different and interesting feature of the NNESTs is that many of them set an example by sharing their L1 and native culture in the class in order to create a space for their students to share sociocultural and L1 backgrounds in turn. From the standpoint of the students, being able to make use of the L1 in their thought processes assists their comprehension of the L2 grammar rules. In addition, it relieves students of their workload and consolidates their previous knowledge with the new learning process (Kern, 1994). To add to that, NNESTs' own English learning experience leads them to have a more thorough understanding of the students' learning process (Lasagabaster & Sierra, 2002). Furthermore, it has to be pointed out that the different language proficiency backgrounds make the NESTs' and NNESTs' insight into students' learning process and ways of communicating with students quite different. For many NESTs, cross-cultural misinterpretations are commonplace during teacher-student interactions (Prodromou, 1992). From that viewpoint, studies such as Modiano's (2005) argue that NNESTs have an advantage over NESTs for being open to multicultural perspectives. In Modiano's (2005) words, NNESTs "have knowledge of linguistic complexities of L1 and the target language in contact," and they are also "well suited to provide students with a pluralistic cultural perspective" (p. 26). This facilitates the NNESTs and their students to understand not only the target culture but also each other better, and though NNESTs are strict with language errors, they are more empathetic towards students' language challenges (Reves and Medgyes, 1994). Due to the similar English language learning background, students may also feel psychologically more attached to NNESTs than NESTs (Medgyes, 1992).

More recently, Kang (2008) examined the measures of proficiency, comprehensibility, instructional competence, and accentedness in NS and NNS of English. In her study, both groups evaluated four-minute segments of fourteen NS and NNS teaching assistants' lectures. It was found that NNS gave significantly stricter scores on ratings of comprehensibility and proficiency, but not on accentedness and instructional competence. It was also observed that the more contact that the NNEST raters came into with NNS learners, the more tolerant they became in rating, which lead to higher accentedness scores.

In sum, NESTs and NNESTs diverge in the way they teach, what they focus on and how they connect to their students because of their language proficiency and backgrounds but both groups have a lot to offer with regard to effective teaching practice.

NEST and NNEST as Raters of Oral Performance

According to Zhang and Elder (2011), studies on the dissimilarities between NESTs and NNESTs as raters are few in number, and they have yielded ambiguous results. Although the results are ambiguous, it is still useful to look at what they are in the light of this current study. In the field of language testing NESTs and NNESTs as raters have been compared in terms of how strict they are, how consistent they are, and whether language background or proficiency have an effect on their rating scores. On the one hand, some studies concluded that NNESTs are tougher raters compared to NESTs (Fayer and Krasinski, 1987; Santos, 1988). On the other hand, there are studies that found the opposite to be true (Barnwell, 1989; Hill, 1996). Moreover, several research studies have resulted in the finding that NNESTs are more consistent as raters than the NESTs (Brown, 1995; Hill, 1996), while Shi (2000) has brought forth reverse results. Over the

years, studies have found many differences. However, according to Zhang and Elder (2011), the reasons behind all the deviating findings remain unexplored.

Although mainly unexplored, there are notable studies where the language backgrounds of the NESTs and NNESTs raters have been analyzed. However, the analyses were found to bear no significant impact on differences in their rating preferences (Brown, 1995; Hamp-Lyons & Zhang, 2001; Kim, 2009b; Shi, 2001). It is interesting to note, though, that even if no real difference could be seen between the mean score of the ratings of the NEST and NNEST groups in the quantitative analysis, when similar studies were done qualitatively, the raters' comments disclosed a correlation with their language backgrounds. Overall, the NEST groups seemed more interested in content over form, while the NNEST groups focused on students' use of grammar critically (Brooks, 2013). Brown's (1995) exploration of NEST and NNEST raters suggests that NNESTs follow rating criteria more explicitly, whereas NESTs are more dependent on intuition. However, in the current study, the NEST participants were found to focus more on grammar than the NNEST participants.

Rossiter (2009) studied differences in NS and NNS evaluations of fluency, again recording no significant differences between the groups. What is more, Rossiter discovered that the majority of the negative remarks made by the NS and NNS raters were concerned with fillers, speech rate, pausing, and self-repetition, while there were comments regarding pronunciation, grammar, and vocabulary. Similarly, in the current study, the NEST and NNEST participants took notice of speech rate and fluency as an area of improvement for the ESL student they assessed.

A research study by Kim (2009a) addresses the question of NESTs and NNESTs comparability in English speaking tests. To find results, she compared and contrasted the

evaluations of twelve NEST raters from Canada with twelve NNEST raters from Korea, using mixed-methods analysis. For data collection, the participants evaluated ten examinees who each performed eight varieties of speaking tasks. The product of every individual task was a one-minute of speech that was rated on a four-point scale ranging from "almost always successful" to "almost always unsuccessful." Analysis of the data demonstrated that the NEST and NNEST raters were consistent with similar scores across tasks, with few exceptions. Conversely, Kim also observed that there was inconsistency in the way NEST and NNEST groups dealt with the rating samples. She noticed that NESTs gave more thorough and elaborate ratings for pronunciation, grammar and accuracy. This was also the case in the current study, where it was noticed that the NESTs provided more descriptive feedback to the student than the NNESTs.

As a follow-up on this finding regarding NESTs raters' elaboration on certain elements of speech, Kim (2009b) took on a subsequent study to scrutinize the variations in NEST and NNEST raters' evaluation of ESL students' speaking performances. The follow-up study shed light on the effect of being a NS rater on score reliability and how such reliability also varies according to the number of NS and NNS raters taking part in the scoring. The current study also found the score reliability to be higher with the NESTs than the NNESTs who were also fewer in number. Therefore, on the basis of Kim's findings regarding how reliability varies according to the number of NEST and NNEST participants, it can be said that if the NNESTs were more in number, the current study would have yielded different results in terms of inter-rater reliability. Kim (2009b) also found that examinees' ability played a bigger role in the deviation of the scores than rater effect. Hence, the rater groups were concluded to show comparable harshness patterns in overall scores.

More recently, a study was done by Zhang & Elder (2011) where the purpose was to look at the NEST and NNEST raters' assessment of the College English Test-Spoken English Test (CET-SET). The study addressed the question of whether the way NNESTs judged language proficiency corresponded with NESTs, or if there were differences. Results turned out to reveal no significant difference in the raters' assessment. In fact, it revealed a general level of agreement between NEST and NNEST raters on the oral construct components. However, the study only shed light on the unguided holistic evaluation of oral performance without the use of any guided rating scale. The current study tries to answer the more important question of how the NEST and NNEST raters apply the different categories on a guided rating scale for oral performance. Its aim is to determine the categories they attach more importance to and the reasons behind that inclination.

Furthermore, an interesting study by Brooks (2013) examined the raters' scoring construct. In her study, 30 NS and NNS Federal Bureau of Investigation (FBI) speaking testers participated in rating 25 English Speaking Proficiency Tests. The rater-participants developed their own rating categories, comprising of functions, organization, structures, vocabulary, fluency, pronunciation and social/cultural appropriateness. The study could not find any statistically significant difference between the NS and NNS testers. However, when raters were divided into groups depending on their English proficiency level, irrespective of their NS or NNS position, it was found that lower proficiency raters gave significantly lower ratings. Therefore, Brooks (2013) pointed how it is necessary for the rater training organizations to take note of rater proficiency level regardless of whether they are NS or NNS.

It is also important to mention that in addition to NEST and NNEST raters' L1 background and language proficiency, studies have also been performed on other rater

background variables such as training (Hill, 1996; Weigle, 1994, 1998), previous English language teaching/learning experience (Cumming, Kantor, & Powers, 2002; Weigle, Boldt, & Valsecchi, 2003), and familiarity with local accents (Carey, Mannell, & Dunn, 2011). All these variables, turned out to be an influence on NEST and NNEST raters' responses to ratings.

To summarize, it can be said that even though early studies that compared the rating of NEST and NNEST rater groups have brought forth mixed results, more recent research studies did not find evidence of significant differences between NEST and NNEST raters (Kim, 2009b; Zhang and Elder, 2011). However, studies have also demonstrated that NESTs as raters seem more focused on communication skills, whereas the NNESTs as raters seem more form-focused. Some of these findings of the comparative and contrastive studies between NEST and NNEST raters are consistent with the current study while some are not. The discussion of the relation between the reviewed studies and the current study can be found in Chapter five.

Issues in the Assessment of Speaking

According to O'Sullivan (2012) speaking tests are the most challenging to design and oversee, and the most notable difficulties are relevant to "the predictability of the task response, the validity and the consistency of the scoring system" (p. 234). This segment of the chapter focuses on the studies conducted in recent years on various aspects of speaking such as the impact of test takers or students on their oral performance, task parameters manipulation, rating scales, and rater performance. For the current study, that uses a rating scale as a data collection instrument, the discussion of having a task appropriate rating scale is important. In addition, understanding of rater performance as an issue in speaking assessment adds to the discussion of inter-rater reliability. Firstly, it has been noted that test taker or students' characteristics such as

language ability, as well as personality, age and gender have an impact on oral performance (Berry, 2004; O'Loughlin, 2001). Secondly, manipulation of task performance conditions or parameters by adding to or reducing from allocated presentation time has effects on the performance (O'Sullivan et al., 2006). Thirdly, the rating scale should be context appropriate and should correspond to the definition of the construct of the performance task (Fulcher, 1996; North, 1995). Fourthly, how the rater performs with the rating scale when scoring and how he/she reacts to the test taker has an influence on the oral performance (O'Sullivan, 2002; O'Sullivan and Rignall, 2007). The following is a brief discussion of each of these four aspects.

Test Takers

As far as research pertaining to the effect of physical characteristics on oral performance goes O'Sullivan (2000) states that variables such as age, gender, perception of language ability and personality may be contributing factors towards inconsistency in the performance of a speaking task. With that being said, it is important to highlight that the evidence of gender as a factor is rather diverse. O'Sullivan's (2000) study resulted in real divergence in test performance where the examiner was female for both male and female test takers, while O'Loughlin's research (2002) concludes that having a female examiner did not have any significant effect on the examiners in the IELTS interviews from the viewpoint of neither the test test-taker nor language and scoring.

Task Performance Parameters

Within the topic of task performance, the most frequently investigated aspect has been the planning of time (Foster & Skehan, 1997). Through established research, it is now well

known that appropriately planned time can result is significant improvement in the performance

of oral tasks (O'Sullivan, 2014). By the same token, other parameters that can benefit from

broader research include how the prior understanding of rating criteria of the performance, and

how the knowledge of the amount of language output expected during the delivery will influence

the test performance (Weir, O'Sullivan, & Horai, 2004). In addition, the mode of delivering the

test to the candidates should also be taken into account. According to O'Sullivan (2014), the

commonly used test formats are live, recorded, and automated. Here only the recorded format of

test taking will be explained as a recorded oral presentation was used in the current study. In the

recorded test format the performance is taped on an audio or video device for later scoring by

human raters using a pre-established rating scale. During the recording, the test taker can

perform the task with an examiner present or he/she could also be responding to previously

recorded, written or visual prompts.

Rating Scale

The major questions connected to any speaking task performance are regarding the

scoring. Out of those major questions, the two questions that are particularly crucial are:

1. How will the performance be scored?

2. What type of scale will be used?

In the current study, participants were asked to score the sample ESL student's oral

presentation using a rating scale, which was a very important instrument in the data collection

process. Isaacs and Thomson (2013) briefly describe rating scales as, "the framework within

which human raters score second-language (L2) performance, which is taken to be an indicator

of L2 learners' ability on the construct being measured" (p. 135). There are mainly two broad

categories from which to choose a task, test taker and rater appropriate rating scale. The categories are, namely, holistic and analytic.

According to O'Sullivan (2012), the holistic scale is the one that has a single all-inclusive grade or score that is based on a sequence of descriptors. When using the holistic scale to rate, the rater or examiner matches the student's oral performance to the closest descriptor. There are several advantages of using holistic rating scales. Its biggest benefits are the simplicity of the scale and the speed it allows in the scoring process. Furthermore, training raters how to come to an agreement within a band is less cumbersome. However, O'Sullivan (2012) explains that the disadvantage of this scale lies in the danger of "trial by first impression" (p. 242). Trial by first impression refers to the fact that to give one score only, the examiner may by practice depend on his/her first impression of the test taker. Secondly, because of being holistic, the bands and the descriptors on the scale may not represent more than a very crude measure for the test taker's abilities.

Respectively, O'Sullivan (2012) also discusses the analytic rating scale. On an analytic type of scale, the language functions of the task in which it will be used are identified in advance. Secondly, a scoring scheme specific to those language functions is created. This method of developing the rating scale makes it multi-faceted with several components that add to an overall score. Interestingly however, the analytic scale has been criticized for being a collection of several holistic scales in one. It has been said that the distinction between the two scales is unclear. Therefore, the shortcomings of the analytic rating scale is comparable to those of the holistic rating scale, expect that in the case of analytic scales, the disadvantages are magnified by the inclusion of a greater number of criteria. Nonetheless, unlike holistic scales, the key

advantage of the analytic scale is that it is regarded as more reliable and accurate for simpler speaking tasks.

For the purpose of the current study, an analytic scale was used to make sure that the descriptors on the scoring scale represented the verbal and non-verbal skills to be measured and judged for the sample student performance. In addition, as analytic scales are considered more reliable and accurate for less complex tasks, it was the best choice for the current study.

Rater Performance

The first point that O'Sullivan (2014) raised in his discussion of rater performance is the importance of rater training prior to the rating process. This is necessary in order to maintain the integrity of the scoring process. According to Lumley and McNamara (1995), the training should focus on things like intra-rater consistency and critical boundary realization. In other words, the raters should be trained to be consistent and be able to automatically recognize a passing or failing performance, and to have high inter-rater reliability. In the training, it is also required to help the raters comprehend the rationale behind the rating scale and task for which they are being trained.

Importance of Oral Presentations

In a study conducted by Zitouni (2013) on the use of the oral presentation to improve university EFL students' oral proficiency, findings show that oral presentations are advantageous for students to enhance their performance in speaking. The study further holds that the acquisition of presentation skills in English classes also helps them in other courses.

An older study by Malik and Gulnaz (2011) shows that oral presentations are beneficial for the improvement of speaking. However, Malik and Gulnaz also draw attention to the fact that oral presentations as a tool can have several drawbacks, particularly in the EFL context. The drawbacks they have mentioned are as follows, " lack of teachers' interest in assigning topics selected by students, non-sympathetic behavior of teachers, lack of guidance from the teachers, inequality in time allocation for each presentation, unsatisfactory evaluation of oral presentations and second language barriers" (Malik and Gulnaz, 2011, p. 343).

In addition, according to a study by Bhati (2012), international students tend to score low on oral presentations especially when it has to be performed as a group work. The international students find it harder to collaborate effectively in groups, and their limitations regarding group work is reflected on their oral presentations.

Therefore, it can be said that even though research shows that oral presentations are beneficial in developing student's speaking skills, there are supporting factors such as the role of the teachers and ability of the students to work in groups, which can have an impact on the success of oral presentations as a tool of assessment.

Conclusion

To sum up, it can be said that a number of research works have been done over the years exploring the differences between NESTs and NNESTs as raters. The more recent among those studies do not note any statistical significance between the consistency of NS and NNS raters. Therefore, this area has great scope for further investigation. Moreover, unlike the current study that explores which assessment criteria on the rating scale the NESTs and NNESTs regard as most or least important in the evaluation of oral presentations, the previous studies did not take

the differences in the preferences of evaluation criteria between NESTs and NNESTs into

consideration.

Chapter III

Methodology

This chapter explains how the present study was conducted in order to find out: 1) if there are any differences between NESTs and NNESTs in the way they assess an US university ESL student's oral presentations, and if so, what those differences are; 2) the inter-rater reliability among all the participants, among the NEST group, and among NNEST group; and 3) the assessment criteria each group thinks are more important or less important in the process. The chapter includes an explanation of each of the following: the research setting and participants, research materials, data collection process, and data analysis.

Research Setting and Participants

The current research was conducted at Minnesota State University, Mankato (MNSU), as a graduate thesis paper, which is a requirement for completion of my Master's degree in TESL. The participants in this study were professors and instructors (including graduate teaching assistants) in the Department of English and the Department of Communication Studies at MNSU. In total there were 31 participants including 12 males and 19 females. Among the 31 participants, there were 19 NESTs with six male and 13 female participants, and there were 12 NNESTs with six male and six female participants. The mean age for the NEST participants was 35.88 and for the NNEST participants was 29.83. Altogether, for all of the participants the mean age was 33.38 (see Table 3.1). All the NEST participants were nationally American except for one who was from Singapore. The nationalities of the NNEST participants were mixed. There was one participant each from the following countries: Nepal, Germany, Ukraine, Uzbekistan,

India, Nigeria, Moldova, Finland, and Japan. There were three participants from Bangladesh.

These participants' L1 were respectively: Nepali, German, Ukrainian, Tajik, Malayalam, Igbo,

Romanian, Swedish, Japanese and Bangla. Out of all of the 31 participants, 21 had the

experience of assessing the speaking of an ESL student before, and 10 had no related experience.

Among the 21 participants with experience, there were 15 NESTS and six NNESTs while there

were four NESTs and six NNESTs with no experience (see Table 3.2).

Table 3.1 *Age of the Participants*

Participants	n	Mean	Std. Deviation
All	29	33.38	11.55
NEST	17	35.88	14.04
NNEST	12	29.83	5.42

n=number of participants who mentioned their age

Table 3.2 *Participants with Experience Assessing International Students Speaking*

	Previous Experience	
Participants	Yes	No
All	21	10
NEST	15	6
NNEST	4	6

Research Materials

The data for the current study was collected using the online survey software Qualtrics

through MNSU licensure. The survey consisted of three sections. Section one comprised of a 12-

mintue video of an ESL student's oral presentation from the ESL 125 course, which teaches the

students advanced listening and speaking skills for academic purposes, and a five point Likert-

style analytic rating scale with 10 assessment criteria. The purpose of the rating scale was to record the participants' scores for the student presenter. The criteria on the rating scale were as follows:

1. Communicates the content and his/her thoughts in a clear and coherent manner.

2. Uses own words and speaks naturally (not in a memorized fashion or reading from a text).

3. Speaks fluently and smoothly without too many pauses, fillers (e.g. 'uh'), and/or hesitations.

4. Speaks in a manner which is not too fast or too slow.

5. Pronunciation of individual words/phrases is clear.

6. Uses variance in intonation patterns (not speaking in a monotone voice).

7. Uses correct grammar forms.

8. Makes good eye contact with the audience when speaking (e.g., ensures wide eye contact with all sides of the audience, not frequently looking at notes, looks at audience rather than projection screen, board, or computer screen).

9. Speaks loudly enough so that all audience members can clearly hear.

10. Effective use of body language (posture and gestures).

Firstly, the participants were asked to watch the video only one time and then to fill out the rating scale that followed. At the end of the rating scale, there was an open-ended question for the participants to answer. The question was as follows: If you were to give feedback to the student, what strengths and areas of improvement would you highlight? Please write your comments as they would be given to the student.

Section two of the Qualtrics survey contained another five point Likert-scale question item form. The purpose of that form was to elicit the participants' preferences and perceptions regarding the assessment criteria of oral presentations. The questions that were on the form were as follows:

1. I give (or would give) a lot of importance to correct pronunciation in comparison to other evaluation criteria.

2. I give (or would give) a lot of importance to correct grammar in comparison to other evaluation criteria.

3. I give (or would give) a lot of importance to fluency (smooth flowing speech without too many pauses, hesitations, and/or filler words such as 'uh') in comparison to other evaluation criteria.

4. I give (or would give) a lot of importance to eye contact in comparison to other evaluation criteria.

5. I give (or would give) a lot of importance to body language (posture and use of gestures) in comparison to other evaluation criteria.

6. I give (or would give) a lot of importance to clarity of content delivery in comparison to other evaluation criteria.

7. I think oral presentations are a good tool to measure the speaking skills of international students who have English as their second language.

Section three of the Qualtrics survey consisted of five biographical questions for the participants to answer about themselves. The questions were asked to gather background information about the participants, and they were as follows:

1. What is your gender?

2. What is your age?

3. What is your nationality?

4. What is your first language?

5. Have you assessed the spoken English of an international student before?

Data Collection Process

After the study received Institutional Review Board (IRB) approval, the names and email addresses of each individual faculty member and instructor of the Department of English and the Department of Communication Studies were obtained from the MNSU website. There was a total of 93 currently teaching staff in the two departments at the time. They were all sent an email invitation to take part in the research survey. Each email contained the URL address to the online Qualtrics survey. In addition to containing the URL address, the email explained the objectives of the research as well as the Qualtrics survey procedures. As the email was an invitation to participate in the current study, it asked for voluntary participation through the acceptance of an informed consent form, which was the introductory page of the Qualtrics online survey package. The instructors who were interested in participating in the research study followed the provided URL address to the Qualtrics survey and completed it. The 31 people who completed all the three sections of the survey were counted as participants in this study. The participants had the liberty to complete the survey at the place and time of their preference and convenience. However, they were requested to take the survey within a week of receiving the invitation email, and they were also advised to score the rating scale immediately after watching the video (section one of the survey).

Data Analysis

During the data analysis process, the data in the survey report on Qualtrics was found already organized in accordance with the three main survey sections described under "Research Materials" above. Qualtrics reported the mean and standard deviation of the combined scores (n=31) given on each of the assessment criteria on the rating scale. It also reported the mean and standard deviation of the combined rating (n=31) on each of the question-items on the form about the most and least important criteria of oral presentations in connection to the perceptions of the participants. It then reported the total count of male and female participants, all the individual ages, nationalities and L1. Finally, Qualtrics reported the count of participants with and without previous experience of assessing spoken English of international students. Qualtrics also compiled all the individual feedback provided to the student in one table.

Quantitative Data Analysis

For further data analysis and descriptive statistics, the quantitative data gathered through the online survey on Qualtrics was transferred into the Statistical Package for the Social Sciences (SPSS) 23. Before entering the data into SPSS, the participants were divided into two groups, namely group one NEST and group two NNEST. The quantitative data for the 19 NESTs were entered followed by the quantitative data for the 12 NNESTs with the descriptive variables for both the groups being age, the evaluating criteria, and the preferences and perceptions of the assessment criteria. The descriptive statistics produced the mean and standard deviation of the scores of each of the variables separately for the NEST group and the NNEST group.

Next, an independent samples t-test was performed in order to identify any significant differences between the NEST group and the NNEST group in relation to their assessment scores

for each criterion, and in relation to their perceived preferences regarding the relative importance of each assessment criteria.

Finally, as one of the research questions focused on the extent of inter-rater reliability, a three-way inter-rater reliability analysis in the form of an intraclass correlation coefficient was carried out in SPSS among all the participants together first, only the NEST group next, and only the NNEST group last.

Qualitative Data Analysis

The qualitative data that was collected in the form of the participants' feedback to the student, showed that overall, the NESTs' feedback were more descriptive and well explained than the NNESTs'. For qualitative analysis, the researcher made an attempt to recognize themes or patters across participants' feedback. In the end, the all feedback was categorized according to the following:

1. The assessment criteria the participants commented on the most and the least;

2. The number of NESTs and NNESTs who took notice of those criteria; and

3. The criteria that were regarded as strengths or pointed out as areas of improvement.

Finally, the results of the qualitative data analysis were compared to the results of the quantitative data analysis.

Chapter IV

Results

This chapter presents the results of the study after collecting and analyzing data from 31 participants. The results are presented focusing on the three research questions that guided the study. The research questions are as follows:

1. Are there any differences between native English speaking (NS) and non-native English speaking (NNS) instructors in the way they assess English as a second language (ESL) students' oral presentations in a US university context? If so, what are those differences?

2. What is the inter-rater reliability among:

   i. All the participants;

   ii. Among NS participants, and

   iii. Among NNS participants?

3. Which assessment criteria does each group think are more important or less important?

Data was analyzed using Qualtrics Survey Software (Qualtrics) and Statistical Package for the Social Science (SPSS) 23. Findings pertaining to research question 1 will be reported first.

Findings Pertaining to Research Question 1

This section elaborates on the results of the data analysis in relation to the first research question of this study. To find out the differences between NESTs and NNESTs in the way they assess ESL students' oral presentations, the scores on the assessment criteria on the rating scale in section one of the online survey were analyzed. Additionally, the feedback provided to the

student at the end of the rating scale was also analyzed for any real pattern of difference. For further understanding tables displaying the statistics are provided below.

Quantitative Data Results

Table 4.1 shows the difference in scores for each assessment criteria on the rating scale for all participants in a descending order. From the table it can be seen that the criteria that received the highest scores on the rating scale are assessment criteria two and four with a mean score of 4.29 for both. The lowest score was received by criterion number seven with a mean score of 3.74. Based on this, it can be said that the best element of the student's oral presentation was her natural speech and her speech rate, whereas, the biggest area of improvement in the students oral performance was her grammar.

Table 4.1 *Assessment Criteria Scores for All Participants*

Assessment Criteria	n	Mean	Std. Deviation
#2 Uses own words and speaks naturally (not in a memorized fashion or reading from a text).	31	4.29	0.81
#4 Speaks in a manner which is not too fast or too slow.	31	4.29	0.68
#1 Communicates the content and his/her thoughts in a clear and coherent manner.	31	4.23	0.61
#8 Makes good eye contact with the audience when speaking (e.g. ensures wide eye contact with all sides of the audience, not frequently looking at notes, looks at audience rather than projection screen, board, or computer screen.	31	4.13	0.79
#6 Uses variance in intonation patterns (not speaking in a monotone voice).	31	4.00	0.76
#3 Speaks fluently and smoothly without too many pauses, fillers (e.g. 'uh'), and/or hesitations.	31	3.87	0.79
#9 Speaks loudly enough so that all audience members can clearly hear.	31	3.84	1.02
#10 Effective use of body language (posture and gestures).	31	3.84	0.88
#5 Pronunciation of individual words/phrases is clear.	31	3.77	0.91
#7 Uses correct grammar forms.	31	3.74	0.72

n=number of participants

After looking at the scores on the assessment criteria of all the participants, it is now important to look separately at the scores that the two groups provided on the rating scale. The next two tables show the scores given by the NEST and NNEST groups separately.

Firstly, Table 4.2 shows that the criterion that received the highest score from the NEST group is evaluation criterion number two (M=4.42), and the lowest score was given to criterion number nine (M=3.63). In other words, according to the NESTs the strongest element of the student's oral presentation was her natural speech, while, her speech volume was the biggest area of improvement.

Table 4.2 *Assessment Criteria Scores for NEST Group*

Assessment Criteria	n	Mean	Std. Deviation
#2 Uses own words and speaks naturally (not in a memorized fashion or reading from a text).	19	4.42	0.09
#4 Speaks in a manner which is not too fast or too slow.	19	4.32	0.67
#1 Communicates the content and his/her thoughts in a clear and coherent manner.	19	4.21	0.63
#8 Makes good eye contact with the audience when speaking (e.g. ensures wide eye contact with all sides of the audience, not frequently looking at notes, looks at audience rather than projection screen, board, or computer screen.	19	4.16	0.76
#6 Uses variance in intonation patterns (not speaking in a monotone voice).	19	4.05	0.85
#3 Speaks fluently and smoothly without too many pauses, fillers (e.g. 'uh'), and/or hesitations.	19	3.84	0.69
#5 Pronunciation of individual words/phrases is clear.	19	3.84	0.96
#7 Uses correct grammar forms.	19	3.79	0.79
#10 Effective use of body language (posture and gestures).	19	3.79	0.92
#9 Speaks loudly enough so that all audience members can clearly hear.	19	3.63	1.16

n=number of participants

Secondly, according to the mean scores displayed in Table 4.3, the criteria that received the highest scores from the NNEST group is evaluation criteria number one and four (M=4.25

for both), and the lowest score was given to criteria number five and seven (M=3.67 for both). In

other words, the mean scores show that according to the NNESTs the strongest element of the

student's oral presentation was her clarity and coherence, as well as her speech rate, while the

biggest area of improvement was her pronunciation and grammar.

Table 4.3 *Assessment Criteria Scores for NNEST Group*

Assessment Criteria	n	Mean	Std. Deviation
#1 Communicates the content and his/her thoughts in a clear and coherent manner.	12	4.25	0.62
#4 Speaks in a manner which is not too fast or too slow.	12	4.25	0.75
#9 Speaks loudly enough so that all audience members can clearly hear.	12	4.17	0.72
#2 Uses own words and speaks naturally (not in a memorized fashion or reading from a text).	12	4.08	0.67
#8 Makes good eye contact with the audience when speaking (e.g. ensures wide eye contact with all sides of the audience, not frequently looking at notes, looks at audience rather than projection screen, board, or computer screen.	12	4.08	0.90
#3 Speaks fluently and smoothly without too many pauses, fillers (e.g. 'uh'), and/or hesitations.	12	3.92	1.00
#6 Uses variance in intonation patterns (not speaking in a monotone voice).	12	3.92	0.67
#10 Effective use of body language (posture and gestures).	12	3.92	0.90
#5 Pronunciation of individual words/phrases is clear.	12	3.67	0.89
#7 Uses correct grammar forms.	12	3.67	0.65

n=number of participants

It is now important to compare the scores for the two groups by viewing the differences

in their mean scores and by looking at the t-test statistics of the scores by all the participants. The

following table displays the differences in means and the results of the independent samples t-

test.

Table 4.4 *Difference in Mean Scores and Assessment Criteria T-Test Results based on Two Groups*

Assessment Criteria	Groups	n	Mean	Std. Deviation	Sig. at p< .05
#1 Communicates the content and his/her thoughts in a clear and coherent manner.	NEST	19	4.21	0.63	NS
	NNEST	12	4.25	0.62	
#2 Uses own words and speaks naturally (not in a memorized fashion or reading from a text).	NEST	19	4.42	0.90	NS
	NNEST	12	4.08	0.67	
#3 Speaks fluently and smoothly without too many pauses, fillers (e.g. 'uh'), and/or hesitations.	NEST	19	3.84	0.69	NS
	NNEST	12	3.92	1.00	
#4 Speaks in a manner which is not too fast or too slow.	NEST	19	4.32	0.67	NS
	NNEST	12	4.25	0.75	
#5 Pronunciation of individual words/phrases is clear.	NEST	19	3.84	0.96	NS
	NNEST	12	3.67	0.89	
#6 Uses variance in intonation patterns (not speaking in a monotone voice).	NEST	19	4.05	0.85	NS
	NNEST	12	3.92	0.67	
#7 Uses correct grammar forms.	NEST	19	3.79	0.79	NS
	NNEST	12	3.67	0.65	
#8 Makes good eye contact with the audience when speaking (e.g. ensures wide eye contact with all sides of the audience, not frequently looking at notes, looks at audience rather than projection screen, board, or computer screen.	NEST	19	4.16	0.76	NS
	NNEST	12	4.08	0.90	
#9 Speaks loudly enough so that all audience members can clearly hear.	NEST	19	3.63	1.16	NS
	NNEST	12	4.17	0.72	
#10 Effective use of body language (posture and gestures).	NEST	19	3.79	0.92	NS
	NNEST	12	3.92	0.90	

n=number of participants; NS = non-significant result

Based on the table, we can see that there is little difference in the mean scores for most of the assessment criteria. There appears, however, to be a noticeable difference in the mean scores for criteria two, naturalness of speech, whereby NESTs scored it more highly, and criteria nine, speech volume, in which NNESTs gave it a higher score. Nonetheless, the independent samples t-test revealed that there were no significant differences in the scores for all of the assessment criteria. This is not surprising given the small sample of participants, but it can be suggested that perhaps with a larger sample the results may differ in future studies. At this point, therefore, it is

necessary to shed light on the qualitative data collected through the open-ended question at the end of the rating scale in section one of the online survey on Qualtrics.

Qualitative Data Result

This segment discusses the analysis of the qualitative data collected through the online survey for this current study. At the end of the provided rating scale, all 31 participants were requested to answer an open ended question. The question was phrased as follows:

If you were to give feedback to the student, what strengths and areas of improvement would you highlight? Please write your comments as they would be given to the student.

Twenty-one out of the 31 participants chose to answer the question, and provided their feedback on the student's performance. Out of the 21 participants who commented, there were 14 NESTs and seven non-native English speaking teachers NNESTs. The criterion on the rating scale that was commented on by the most number of participants was 'Uses own words and speaks naturally – not in a memorized fashion or reading from a text.' In total, 11 participants commented on it, including nine NESTs and two NNESTs. Out of the nine NESTs, seven highlighted it as the student's strength and two regarded it as an area of improvement. On the other hand, between the two NNEST participants, one saw it as strength and the other saw it as an area of improvement. It has to be mentioned here that the NESTs high number of comments on 'speaks naturally' supports the mean score in Table 4.4 which was higher than the score for the NNESTs. This shows that the NESTs not only scored it higher but also noticed it more than the NNESTs.

The second most commented on criterion was 'effective use of body language – posture and gestures.' In total, nine participants commented on it, out of which six were NESTs and

three were NNESTs. All the participants regarded it as an area of improvement, except for one NEST participant who stated that the student "used an appropriate amount of natural gestures."

Next, a total of eight participants including six who were NESTs and two who were NNESTs commented on 'communicates the content and her thoughts in a clear and coherent manner.' Five participants, four NESTs and one NNEST regarded it as the student's strength, whereas three participants, two NESTs and one NNEST, regarded it as an area of improvement. A total of eight participants also commented on 'makes good eye contact with the audience when speaking (e.g. ensures wide eye contact with all sides of the audience, not frequently looking at notes, looks at audience rather than projection screen, board, or computer screen).' Out of the eight, seven were NESTs and only one was NNEST. It was regarded as strength by three NESTs and as an area of improvement by four NESTs and one NNEST.

In addition, seven participants commented on 'speaks loudly enough so that all audience members can clearly hear.' The seven included five NESTs and two NNESTs. All participants indicated it as an area of improvement, except for one NNEST who said, "your voice is clear…." Six participants including four NESTs and two NNESTs commented on 'uses variance in intonation patterns (not speaking in a monotone voice).' All participants regarded it as an area of improvement, except for one NNEST participant, who thought the student, "used correct intonation patterns." Then five participants commented on 'speaks in a manner which is not too fast or too slow.' Out of the five, three were NESTs and two were NNESTs, and all of them unanimously stated that it is an area of improvement, pointing out that she spoke too fast.

Next, four participants commented on 'speaks fluently and smoothly without too many pauses, fillers (e.g. 'uh'), and/or hesitations.' All four were NESTs, which means that no NNEST participants commented on fluency. Three out of the four NEST participants who commented on

fluency identified it as an area of improvement and one NEST participant regarded it as strength and stated, "You were able to speak quite fluently without seeming like your presentation was memorized." There were also four participants who commented on 'pronunciation of individual words/phrases is clear.' Out of the four only one was a NEST and the rest of the three were NNESTs. The one NEST and two NNESTs regarded pronunciation as an area of improvement, whereas one NNEST saw it as strength and stated, "spoke…with good pronunciation."

Finally, a total of four participants commented on grammar, and all of them were NESTs. No NNEST participants provided any feedback on grammar. All four NEST participants consistently indicated grammar as an area of improvement.

From the analysis it is clear that there is an overall inconsistency in the comments of the participants. However, interestingly, none of the NNEST participants commented on fluency and grammar. Conversely, more NNEST participants commented on pronunciation than NEST participants. Tables 4.5 and 4.6 represent the data discussed above. Specifically, Table 4.5 shows the number of NEST and NNEST participants who provided feedback on the student's oral presentation, and on which assessment criteria their feedback was. Table 4.6 displays how many of those comments focused on each assessment criterion as a point of strength (positive feedback) or an area of improvement (negative feedback) for the student. The tables are as follows:

Table 4.5 *Number of Participants who Provided Feedback*

					Assessment Criteria					
n	Clarity of content	Natural speech	Fluency	Speech rate	Pronunciation	Intonation	Grammar	Eye contact	Volume	Body language
NEST	06	09	04	03	01	04	04	07	05	06
NNEST	02	02	00	02	03	02	00	01	02	03
Total	08	11	04	05	04	06	04	08	07	09

n=number of participants who provided feedback

Table 4.6 *Number of Positive and Negative Feedback*

																			Assessment Criteria	
	Clarity of content		Natural speech		Fluency		Speech rate		Pronunciation		Intonation		Grammar		Eye contact		Volume		Body language	
n	PF	NF	PF	NF	PF	NF	PF	NF	PF	NF	PF	NF	PF	NF	PF	NF	PF	NF	PF	NF
NEST	04	02	07	02	01	03	00	03	00	01	00	04	00	04	03	04	00	05	01	05
NNEST	01	01	01	01	00	00	00	02	01	02	01	01	00	00	00	01	01	01	00	03
Total	05	03	08	03	01	03	00	05	01	03	01	05	00	04	03	05	01	06	01	08

n=number of participants who provided feedback; PF=positive feedback; NF=negative feedback

Findings Pertaining to Research Question 2

This section presents the results of the data analysis in relevance to the second research question of this study. The second question relates to the inter-rater reliability among all the participants, among the NEST group, and among the NNEST group. The Intraclass Correlation Coefficient on SPSS was used to calculate inter-rater reliability, which represents the degree of consistency of the assessment process among the raters. When evaluating the Intraclass Correlation, the average measure needs to be greater than 0.70 to be considered acceptable (Larsen-Hall, 2009). The Intraclass Correlation Coefficient for all participants was 0.72; for the NEST group 0.71, and for the NNEST group 0.19. Therefore, it can be said that the inter-rater reliability among all the participants and separately among the NEST group is good. However, the inter-rater reliability among the NNEST group was quite low.

Findings Pertaining to Research Question 3

The final segment of this chapter focuses on the data in relation to the third and last research question of this study. The third research question investigates which assessment criteria of the oral presentation are most or least important to the participants in relation to assessing the student's oral presentation performance. Similar to the data analysis of the difference in scores by the NESTs and NNESTs, the mean and standard deviation scores were calculated on SPSS to identify any differences. Next, a t-test was carried out to disclose any significant difference in the data. For a clearer understanding, tables displaying the statistics are provided below.

Table 4.7 shows the perceptions of criteria importance scores for all participants. From the table it can be seen that the question-item with the highest mean score across all participants was the one pertaining to the preference of giving 'a lot of importance to clarity of content

delivery' (M=4.48). The question-item that received the lowest score pertains to giving 'a lot of importance to correct pronunciation' (M=2.71).

Table 4.7 *Perceptions of Criteria Importance Scores for All Participants*

Question Items	n	Mean	Std. Deviation
I give (or would give) a lot of importance to clarity of content delivery in comparison to other evaluation criteria.	31	4.48	0.84
I think oral presentations are a good tool to measure the speaking skills of international students who have English as their second language.	31	3.94	0.95
I give (or would give) a lot of importance to eye contact in comparison to other evaluation criteria.	31	3.84	1.08
I give (or would give) a lot of importance to fluency (smooth flowing speech without too many pauses, hesitations, and/or filler words such as 'uh') in comparison to other evaluation criteria.	31	3.52	1.04
I give (or would give) a lot of importance to body language (posture and use of gestures) in comparison to other evaluation criteria.	31	3.48	1.10
I give (or would give) a lot of importance to correct grammar in comparison to other evaluation criteria.	31	2.90	1.17
I give (or would give) a lot of importance to correct pronunciation in comparison to other evaluation criteria.	31	2.71	1.25

n=number of participants

After looking at the mean scores for all the participants, it is now important to look separately at the mean scores of the two groups. The next two tables show the means scores for the NEST and the NNEST groups.

Table 4.8 shows the perceptions of criteria importance scores for the NEST participants. It reflects the same pattern demonstrated by all the participants with 'clarity of content' as the highest mean (M=4.32) and 'pronunciation' as the lowest mean (M=2.42).

Table 4.8 *Perceptions of Criteria Importance Scores for NEST Group*

Question Items	n	Mean	Std. Deviation
I give (or would give) a lot of importance to clarity of content delivery in comparison to other evaluation criteria.	19	4.32	1.00
I give (or would give) a lot of importance to eye contact in comparison to other evaluation criteria.	19	3.68	1.16
I think oral presentations are a good tool to measure the speaking skills of international students who have English as their second language.	19	3.63	0.96
I give (or would give) a lot of importance to fluency (smooth flowing speech without too many pauses, hesitations, and/or filler words such as 'uh') in comparison to other evaluation criteria.	19	3.37	1.01
I give (or would give) a lot of importance to body language (posture and use of gestures) in comparison to other evaluation criteria.	19	3.37	1.12
I give (or would give) a lot of importance to correct grammar in comparison to other evaluation criteria.	19	2.89	1.37
I give (or would give) a lot of importance to correct pronunciation in comparison to other evaluation criteria.	19	2.42	1.30

n=number of participants

Next, Table 4.9 shows the perceptions of criteria importance scores for the NNEST participants. From the table it can again be seen that the question-item with the highest mean score among the NNESTs is the one pertaining to the preference of giving 'a lot of importance to clarity of content' (M=4.75). The question-item that received the lowest score pertains to giving 'a lot of importance to correct grammar' (M=2.92).

Table 4.9 *Perceptions of Criteria Importance Scores for NNEST Group*

Question Items	n	Mean	Std. Deviation
I give (or would give) a lot of importance to clarity of content delivery in comparison to other evaluation criteria.	12	4.75	0.45
I think oral presentations are a good tool to measure the speaking skills of international students who have English as their second language.	12	4.42	0.79
I give (or would give) a lot of importance to eye contact in comparison to other evaluation criteria.	12	4.08	1.00
I give (or would give) a lot of importance to fluency (smooth flowing speech without too many pauses, hesitations, and/or filler words such as 'uh') in comparison to other evaluation criteria.	12	3.75	1.14
I give (or would give) a lot of importance to body language (posture and use of gestures) in comparison to other evaluation criteria.	12	3.67	1.15
I give (or would give) a lot of importance to correct pronunciation in comparison to other evaluation criteria.	12	3.17	1.11
I give (or would give) a lot of importance to correct grammar in comparison to other evaluation criteria.	12	2.92	0.90

n=number of participants

It is now essential to compare the scores for the two groups by viewing the differences in their mean scores and by looking at the t-test statistics. Table 4.10 displays the differences in means scores between the two groups and the results of the independent samples t-test.

Table 4.10 *Difference in Mean Scores and Perceptions of Criteria Importance T-Test Results based on the Two Groups*

Question Item	Groups	n	Mean	Std. Deviation	Sig. at p< .05
I give (or would give) a lot of importance to correct pronunciation in comparison to other evaluation criteria.	NEST NNEST	19 12	2.42 3.17	1.30 1.11	NS
I give (or would give) a lot of importance to correct grammar in comparison to other evaluation criteria.	NEST NNEST	19 12	2.89 2.92	1.37 0.90	NS
I give (or would give) a lot of importance to fluency (smooth flowing speech without too many pauses, hesitations, and/or filler words such as 'uh') in comparison to other evaluation criteria.	NEST NNEST	19 12	3.37 3.75	1.01 1.14	NS
I give (or would give) a lot of importance to eye contact in comparison to other evaluation criteria.	NEST NNEST	19 12	3.68 4.08	1.16 1.00	NS
I give (or would give) a lot of importance to body language (posture and use of gestures) in comparison to other evaluation criteria.	NEST NNEST	19 12	3.37 3.67	1.12 1.15	NS
I give (or would give) a lot of importance to clarity of content delivery in comparison to other evaluation criteria.	NEST NNEST	19 12	4.32 4.75	1.00 0.45	NS
I think oral presentations are a good tool to measure the speaking skills of international students who have English as their second language.	NEST NNEST	19 12	3.63 4.42	0.96 0.79	-2.37

n=number of participants; NS = non-significant result

Even though there appears to be a noticeable difference in the mean scores for question-items about 'pronunciation', 'fluency' and 'clarity of content', whereby NNESTs scored them more highly, the independent samples t-test did not reveal that there were any significant differences between them. Of all the question items, only the one related to the perception of viewing 'oral presentations as a good tool to measure speaking kills' was significant at p < .05 with NNESTs viewing them as more important compared to NESTs.

Chapter V

Discussion and Conclusion

This chapter discusses the results of the data analysis based on the three research questions of this study. The research questions are as follows:

1. Are there any differences between native English speaking (NS) and non-native English speaking (NNS) instructors in the way they assess English as a second language (ESL) students' oral presentations in a US university context? If so, what are those differences?

2. What is the inter-rater reliability among:

   i.   All the participants;

   ii.   Among NS participants, and

   iii. Among NNS participants?

3. Which assessment criteria does each group think are more important or less important?

The following segments look over the findings related to the research questions in light of previous research studies on NESTs' and NNESTs' assessment of ESL students' oral presentations, and concludes with the limitations of the study, suggestions for future research, and the researcher's observations and pedagogical implications.

Differences between NESTs and NNESTs in the way they assess English as a second language (ESL) students' oral presentations

The interpretation of the data regarding the different assessment criteria on the rating scale for all participants demonstrated that the best element of the student's oral presentation was her natural speech and her speech rate, whereas, the biggest area of improvement was grammar.

When looking at the NESTs and NNESTs separately, it was seen that according to NESTs, the strongest element of the student's oral presentation was her natural speech, while, speech volume was the biggest area of improvement. The NNESTs' scores revealed that the strongest element of the student's oral presentation according to them was clarity and coherence, as well as speech rate, while the biggest area of improvement in the student's oral presentation was her pronunciation and grammar. Even with these inconstancies, no statistically significant difference was found in the assessment scores of NESTs and NNESTs.

However, it is crucial to discuss the possible reasons behind these inconsistencies in light of previous research studies. According to previous studies, NESTs have more language authenticity, for example their pronunciation is accurate, their range of vocabulary is wide and they have full cultural understanding of the language (Baratt & Kontra, 2000). Comparatively, NNESTs may be less competent speakers, with language that sounds bookish (Medgyes, 1994). In Braine's (2005) opinion, the majority of the different teaching practices employed by NESTs and NNESTs correlate to their differing language proficiency levels. Although the current study did not gather data regarding the language background to be able to draw a valid conclusion, the difference in language proficiencies could be considered as the reason why during assessment, the NESTs took more notice of how naturally the student spoke while the NNESTs paid more attention to the coherence and grammatical accuracy of the speech.

Subsequently, it is important to compare and contrast the result of the quantitative data with the result of the qualitative data. The analysis of the qualitative data revealed that a high number of NESTs commented on the assessment criterion regarding natural speech and most of the comments were positive. This supports the noticeable difference in mean scores between the two groups about the natural speech criterion. Even though that difference did not turn out to be

significant, it is clear enough that the NESTs noticed naturalness of speech more and gave it a higher score compared to the NNESTs. On the other hand, most NESTs provided negative comments on 'speech volume.' This result is inconsistent with the quantitative result as the highest mean score in NEST group was received by natural speech and the lowest was received by speech volume.

It has been mentioned earlier that the difference in language proficiency of the NESTs and NNESTs might possibly have played a role in the noticeable inconsistency in their scores for natural speech. Similarly, the inconsistency in their scores for speech volume could be related to the results from previous studies that held that that NESTs have the tendency to exercise greater leniency towards student's language errors as long as communication is effective while NNESTs are more inclined to focus on and give feedback to students' language errors (Lasagabaster & Sierra, 2002; Medgyes, 2001). From here, it can be understood that perhaps because NESTs tend to ensure that communication is not being hindered, they focused more on the student's volume of speech than the NNESTs did.

In the NNEST group, a high number of participants commented on 'pronunciation' and 'body language' and most of those comments were negative. A low number of NNEST participants commented on 'eye contact' and that comment was negative. This result is partly consistent with the quantitative data because according the lowest mean score in the NNEST group was given to 'pronunciation' on the rating scale.

In addition, the qualitative data also found that interestingly, none of the NNEST participants commented on fluency and grammar if though they mean score of grammar was low. Conversely, more NNEST participants commented on pronunciation than NEST participants.

If this finding is discussed in the light of the literature review, it can be said that the current study supports the claim that NNESTs are tougher raters compared to NESTs as some studies have indicated (Fayer and Krasinski, 1987; Santos, 1988) because the NNESTs' mean scores were lower than NESTs' mean scores on six out of four assessment criteria on the rating scale (see Table 4.4). However, it does not support the conclusion that the NEST groups are more interested in content over form or that the NNEST groups assess grammar very critically (Brooks, 2013) because in this study, the NNESTs scored the content criteria more liberally than NESTs, and none of the NNESTs provided corrective feedback on grammar use.

It is also important to mention that on average the feedback provided by the NESTs were more descriptive and well explained. An example of feedback provided by the NEST participants is as follows:

You are an effective communicator and you got more and more comfortable as the presentation went on. There is not one thing that gets in the way of your comprehensibility. A few things could make you an even better presenter: 1. Think about the use of your body. How might you use gestures and facial expressions to help make your presentation a bit more dynamic? Your warmed up and got much more natural by the end of the presentation, so you might consider how to do that sooner. 2. How cud you use stress intonation, and volume to vary your speech a bit more and emphasize key ideas? 3. You were able to speak quite fluently without seeming like your presentation was memorized. Sometimes it's hard to balance this fluency with accuracy. And while the small grammatical errors did not impede comprehensibility, I wonder how you might be able to give a bit more attention to accuracy. Overall, great job!

On the other hand, an example of feedback provided by the NNEST participants is as follows:

Good job! One of the key strengths of her presentation is your confidence. You seem pretty confident throughout the presentation with a smiley face. You explained everything nicely. The thing I'd to suggest you for the future reference is the pace of your presentation and your eye contact with the audience. Sometimes you seem in a rush little bit. And having balanced eye contact toward all audience would be better.

This can be supported by Kim (2009b) who stated that the qualitative analysis of her study demonstrated that the judgements of the NESTs were more detailed and elaborate than those of NNESTs in the areas of pronunciation, specific grammar use and the accuracy of transferred information.

Inter-rater reliability among all the participants, among NS, and among NNS participants

The results of the second question, which looked to determine the inter-rater reliability among all the participants, among NEST, and among NNEST participants showed that the Intraclass Correlation Coefficient for all participants is 0.72, for NEST participants 0.71 and for NNEST participants 0.19. It can be concluded that the inter-rater reliability among all the participants and separately among the NEST participants as a group is good. However, the inter-rater reliability among the NNEST participants as a group is quite low.

Connecting back to the review of the literature, it can be said that the reliability test results do not support some of the previous studies such as Brown (1996) and Hill (1996) where the NNESTs are more consistent as raters than the NESTs, while the results support others such as Shi (2000) where NESTs are more consistent. A study by Kim (2009a) shed light on the effect

of being a NS on score reliability and how such reliability also varies according to the number of NS and NNS raters taking part in the scoring. In the current study the number of NESTs was greater than the number of NNESTs. A larger group of NNEST participants could have yielded a more consistent result.

Also the amount of training that raters receive has a connection to inter-rater reliability (Lumley and McNamara, 1995). While the current study provides information about previous ESL speaking assessment experience of the participants (21 with experience; 10 without experience), from that data there is no way to decide how many of the participants were professionally trained in rating oral presentations.

Most and Least Important Assessment Criteria in Participants' Perceptions

Although there are some noticeable differences in relation to the importance of pronunciation as an assessment criterion, where NNESTs viewed it as more important than NESTs, the difference did not have statistical significance. However, it is still interesting because according to previous studies, due to the effect of language proficiency, NESTs tend to give more importance to pronunciation than NNESTs (Celce-Murcia & Larsen-Freeman, 1999). If the noticeable difference in the importance of pronunciation were statistically significant, then it would have yielded a result different from the previous studies.

A significant difference, meanwhile, was found in the degree to which NESTs and NNESTs perceive oral presentations as a good assessment tool to measure the speaking skills of international ESL students. Between the NEST and the NNEST group of participants, the NNEST group viewed it as a more important tool than did the NEST group. There has not been too many recent studies done on the importance of oral presentations in enhancing speaking

proficiency, but according to the studies by Malik and Gulnaz, (2011) and Zitouni (2013), both students and teachers found that oral presentation successfully help students to improve their spoken skills. Additionally, Bhati (2011) and Malik and Gulnaz (2011), shed light on the existence of supporting factors such as the role of the teachers and the ability of the students to work in groups, which can have an impact on the success of oral presentations as a good assessment tool. Therefore, perhaps due to the challenges that make the planning and administering of oral presentations difficult (O'Sullivan, 2012), the NESTs do not perceive it as a good assessment tool. Again, this study does not have concrete data to establish any reason behind the significant difference pertaining to the data analysis of the most and least important assessment criteria in participants' perceptions.

In the end, it is essential to mention that recent studies of comparison between various aspects, which can influence the assessment practices of NESTs and NNESTs, have not been able to produce any statistically real difference (Zhang & Elder, 2011). This is clearly also the case within the present study. Nonetheless, the main highlight of this study is that unlike any other previous study, it produced a significant difference in the way oral presentations are perceived as good assessment tool to measure speaking ability with NNESTs thinking more highly of it compared to the NESTs.

Limitations

The first and foremost limitation of this study is the limited number of participants. Surely, a larger sample would have yielded different results. A larger population of participants including faculty member and instructors from other universities may have produced significant differences in the assessment process and inter-rater reliability across all groups.

The second limitation renders to not having included more biographic questions concerning the participants' education and professional back ground. Questions such as the following would have yielded more depth to the data:

i.   For how many years have you studied English?

ii.   Did you study English as a foreign language or a second language?

iii.   What aspects of the English language do you think you are more skilled at?

iv.   Have you had professional training on how to assess learners' speaking skills?

Answers to these questions could have provided more qualitative data to compare the effects of language background, proficiency and training on differences in the assessment practices between NESTs and NNESTs and their inter-rater reliability.

The third limitation is related to the use of questionnaires to collect data. Generally, questionnaires have some limitations that can lead to the production of unreliable and invalid data (Dornyei, 2010). Two of the main problem sources of questionnaires are as follows:

1) They yield simple and superficial answers.

As most questionnaires are self-completed, the questions need to be simple and straightforward to avoid confusions and misunderstandings. However, the simplicity of the questions is a limitation because then it becomes impossible to elicit insightful information through the questionnaire.

2) Social desirability bias

People have a natural tendency to represent themselves in good light. Therefore, more often than not, they provide desirable, acceptable and expected answers to the researcher's questions in the questionnaire instead of providing honest answers. As there are no foolproof

ways of ensuring that no false answers are being given, it is a shortcoming of using

questionnaires as a data collection instrument. Having a multiple variety of data collection

instruments such as interviews could have elicited more in depth information. Here it is

important to address the reasons this study was unable to include multiple data source. This study

prioritized the simplicity and swiftness of the data collection process to ensure maximum

participation and completion of the online survey from the prospective participants. One on one

interviews would have been time consuming and for working teachers, making sufficient time

available from their schedules may not have been possible. Moreover, as the participants in the

current study were from among the teachers and colleagues of the researcher, it was assumed that

they may choose not to answer too many biographical questions even after taking all the

necessary steps to keep the data anonymous. For that reason, additional biographic questions

regarding the participants' educational and professional background were not added in this study.

For future research it is important to involve more participants from different universities to

avoid such issues.

Suggestions for Future Research

It has already been mentioned that a larger population of NESTs and NNESTs has to be

approached, preferably at various other colleges and universities to participate in the data

collection process of future, similar studies. Additionally, more open-ended questions to elicit

elaborate qualitative data regarding the participants' language learning, proficiency and

professional backgrounds should be added to the survey.

Moreover, it is important to add interviews to the data collection process by having

interviews with a select number of participants from both groups after they have filled out the

questionnaire to get deeper clarifications regarding their answers on the questionnaire. For instance, in a future study, it will be useful to find out the reasons why participants rate in a particular way, and why they view some criteria as being more important than others. The only way to really do this effectively is through interviews. Further studies can also look at the specific interesting results in the current study in relation to the importance of oral presentations as an assessment tool, the result about naturalness of speech and the difference between the two groups, and the result about the importance of pronunciation. Also, the finding that no NNEST participants provided feedback or took notice of speech volume as an area for improvement highlights another area for future investigation. Another possibility for future studies can be the investigation of NESTs' and NNESTs' speaking assessment of adolescent (pre-university) learners to compare it with the results of this study done with a university student. The study can be enhanced and enriched by the incorporation of the abovementioned suggestions, and can be expected to yield valid results.

Researchers Observations and Pedagogical Implications

As a GTA myself I had the opportunity to teach ESL 125: Advanced Oracy Skills for Non-native Speakers. It has been mentioned in Chapter two that the video-recording of the oral presentation assessed by the participants was from the ESL 125 course, which teaches the students advanced listening and speaking skills for academic purposes. The aim of this course is to help ESL students improve their academic listening and speaking skills in order to achieve success at an American university. One of the main purposes of oral presentations in this course is to teach students to use different varieties of verbal and non-verbal methods of delivery in a speech. In my opinion, a good oral presentation should have above average performance in all

the assessment criteria present on the rating scale. However, owing to reasons such as language background, teaching preferences and personal bias, some criteria are often preferred over another, or paid more attention to than the next during the assessment process. When I taught ESL 125, the aspect of oral presentations that I always noticed first was natural speech. Generally, natural speeches are harder for ESL 125 students to deliver. As international students they are more comfortable when they can read from notes or deliver memorized speeches. Therefore, I always took close notice when a student's speech was natural. Also, I would provide the student with sufficient positive feedback to encourage the continuity of the practice of delivering speeches naturally. Basing on my personal experience, I can assume that in this study the NESTs took more notice of natural speech because they probably were more accustomed to students reading from notes or delivering memorized speeches.

On the other hand, I feel that as a NNEST and an international GTA, I am more aware of the cultural nuances of the international students. Prodromou (1992) states that for many NESTs, cross-cultural misinterpretations are commonplace during teacher-student interactions. On the contrary, it is argued that NNESTs have an advantage over NESTs for being open to multicultural perspectives (Modiano, 2005). In other words, while NESTs are more adept to provide the best example of the ELLs' target language culture, NNESTs are more empathetic towards the learners' sociolinguistic challenges. With that in mind, I can formulate that the NNESTs did not focus on the speech volume because they empathized with the possibility that in the culture to which the ESL student belonged, people were traditionally soft spoken as speaking loudly was not socially encouraged.

Having said that, I would like to mention that L2 acculturation is crucial in the SLA process of ELLs. To make that process successful, I suggest that the NNESTs help the ESL

students by using their empathy to bring the students out of their comfort zone and reduce their affective filters.   In my opinion, this can successfully be done by guiding the students to let their L1 background assist their SLA process. For instances, both NNESTs and NESTs can allows the students to use their L1 in the initial brainstorming and preparation stages of their oral presentations to ease their thinking process and encourage the generation of better ideas.

I would now like to shed light on the result which showed that that there was 0% feedback from the NNESTs on grammar even if they scored the student low in it. As it has been mentioned in the introduction of this study, there is a rapid growth in the adoption of CLT in the ESL context. The incorporation of CLT leaves minimal scope to focus on grammar, and as NNESTs by way of their educational background are more comfortable to teach grammar rules, they usually undergo training to focus more on meaning than form in spoken language. It is possible that the NNESTs in the current study were consciously avoiding the feedback on grammar because they paid more attention to clarity and coherence of the content. Nevertheless, it is imperative that the NNESTs know the appropriate ways in which to provide constructive feedback on grammar mistakes without making the importance of accuracy paramount. Additionally, NNESTs can incorporate mini language lessons in oracy classes to highlight specific linguistic areas of improvement for students.

Furthermore, I would like to discuss about inter-rater reliability. According to O'Sullivan (2014), rater training prior to the rating process is very important. Rater training maintains the integrity of the scoring process. In the current study, the participants were not trained to use the rating scale. Even though most of them had previous related experience, it is possible that they were accustomed to use a rating scale very different from the one used in this study. If the participants could be trained to use the rating scale before taking the online survey, their scores

could have been more reliable. Therefore, this implies that context specific training regarding the use of rating scales and providing feedback is essential especially for the NNESTs who rate speaking skills of university ESL students, in order to maintain high inter-rater reliability among the raters' assessment of oral presentations.

Finally, to discuss about the significant difference found in NESTs and NNESTs preference for oral presentations, I can say that for international students oral presentation is a new and interesting alternative to traditional assessment. Most NNS individuals, especially from the EFL context, are more familiar with the traditional methods of assessment such as tests and exams as oral presentations were not a common part of the language learning process in their schools and universities. In some cultures, oral presentations are still not as widely used as in the US universities. The NNESTs working in the ESL context have the opportunity to compare the benefits of having and not having oral presentations as part of the language learning process. That is the reason why they can acknowledge the importance of oral presentations more. I believe the NNESTs in the current study value oral presentations as important because they are aware of the advantages of oral presentations both as users and teachers of the English language.

Summary

As the global popularity of English learning increases, so does the importance of the NNS group of language teachers. NNESTs have been the focus of many studies since the early 1990s. The differences between NESTs and NNESTs have been studied under a variety of perspectives starting from language proficiency, teaching practices and areas of strengths and weaknesses. Some studies have claimed that in comparison to NESTs, NNESTs have weaker language authenticity and communication skills, but stronger knowledge of grammar and form.

These differences in language proficiency are sometimes said to make way for different teaching perceptions and practices between NESTs and NNESTs. Most of the early studies have been conducted in the EFL setting, and the scope of NNESTs in the ESL environment is rather unexplored by researchers. The few recent studies that have examined NESTs and NNESTs in the ESL environment have not been able to produce any significant differences. With an attempt to explore that less investigated area, this study looked for significant differences in the assessment of speaking between NESTs and NNESTs, their inter-rater reliability, and perceptions regarding the most and least important assessment criteria in oral presentations. After reviewing the literature, it was noticed that many of the studies had their rater-participants assess the speaking of groups of students or sets of tests live. This introduced the extrinsic factors such as test takers, test environment, test administration into the data collection process, and created the possibility of having test validity and content reliability affected by those factors. In the current study, there were no real possibilities for extrinsic factors to influence the data collection process. It was assumed that the unique data collection setting and methods would produce different results. However, quite similar to majority of the recent studies the current study did not reveal any significant differences between the NESTs and NNESTs in their assessment of speaking. However, unlike any other previous study, this study concluded that there is a significant difference in the way NESTs and NNEST perceive oral presentations as a tool for assessing speaking. This also makes the need for context specific training for NNESTs who rate university ESL students' oral presentations more profound because through professional development trainings on how to be consistent in the use of rating scales and providing feedback, raters can improve inter-rater reliability in their speaking assessment practices.

References

Baratt, L., & Kontra, E. H. (2000). Native-English-speaking teachers in cultures other than their own. *TESOL Journal, 9*(3), 19-23.

Barnwell, D. (1989). Proficiency and the native speaker. *Association of Departments of Foreign Languages (ADFL) Bulletin, 20*(2).

Berry, V. (2004). *A study of interaction between individual personality differences and oral performance test facets*. Doctoral theses. London: The University of London.

Bhati, S. (2012). The effectiveness of oral presentation assessment in a Finance subject: An empirical examination. *Journal of University Teaching & Learning Practice, 9*(2).

Boyle, J. (1997). Native-speaker teachers of English in Hong Kong. *Language and Education, 11*(3), 163-81.

Braine, G. (2005). A history of research on non-native speaker English teachers. In E. Llurda, (Ed.), *Non-native language teachers: perception, challenges and contributions to the profession* (pp. 13-23). New York, NY: Springer Science+Business Media, Inc.

Brandl, K. (2008). *Communicative language teaching in action: Putting principles to work*. New Jersey: Prentice Hall.

Brooks, R. (2013). *Comparing native and non-native raters of US Federal Government Speaking Test.* Doctoral thesis. Washington, DC: Georgetown University Press.

Brown, A. (1995). The effect of rater variables in the development of an occupation-specific language performance test. *Language Testing, 12*(1), 1–15. doi:10.1177/026553229501200101

Brown, H. D. (2007). *Teaching by principles: An interactive approach to language pedagogy.* (3rd ed.). New York: Pearson Education.

Canagarajah, A. (1999). Interrogating the 'native speaker fallacy': non-linguistic roots, non-pedagogical results. In G. Braine, (Ed.), *Non-Native Educators in English Language Teaching* (pp. 77-92). Mahwah, NJ: Lawrence Erlbaum.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics, 1*(1), 1-47.

Celce-Murcia, M., Dörnyei, Z., & Thurrell, S. (1995). Communicative competence: A pedagogically motivated model with content specifications. *Issues in Applied Linguistics, 6*(2), 5-35.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). MA, Boston: Heinle & Heinle Publishers.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Crystal, D. (1997). *English as a global language.* Cambridge: Cambridge University Press.

Dornyei, Z., & Taguchi, T. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). London: Routledge.

Eckes, T. (2008). Rater types in writing performance assessments: A classification approach to rater variability. *Language Testing, 25*(2), 155–185. doi:10.1177/0265532207086780

Fayer, J. M., & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37*(3), 313–326. doi:10.1111/j.1467-1770.1987.tb00573.x

Finegan, E. (2004). *Language: its structure and use.* MS, Boston: Thomson Higher Education.

Foster, P., & Skehan, P. (1997). The influence of planning time and task type on second language performance. *Studies in Second Language Acquisition, 18*, 299–323.

Freeman, Y., & Freeman, D. (2003). Struggling English language learners: *Keys for academic success. TESOL Journal, 12* (3), 5-10.

Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing, 13*, 208–238.

Galloway, V. B. (1980). Perceptions of the communicative efforts of American students of Spanish. *The Modern Language Journal, 64*(4), 428–433.

Hamp-Lyons L, Zhang BW (2001). World Englishes: Issues in and from academic writing assessment. In J. Flowerdew, and M. Peacock, (Eds.), *Research perspectives on English for academic purposes.* Cambridge: Cambridge University Press.

Hill, K. (1996). Who should be the judge? The use of non-native speakers as raters on a test of English as an international language. *Melbourne Papers in Language Testing, 5*(2), 29-50.

Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal, 5*(1), 8–11.

Hymes, D. H. (1972). On communicative competence. In J. Pride, & J. Holmes, (Eds.), *Sociolinguistics* (pp. 269-293). Harmondsworth: Penguin.

Institute on International Education (2014). Open Doors 2014: International Students in the United States and Study Abroad by American Students are at All-Time High. [Report]. http://www.iie.org/Who-We-Are/News-and-Events/Press-Center/Press-Releases/2014/2014-11-17-Open-Doors-Data

Joughin, G. (2010). The hidden curriculum revisited: A critical review of research into the influence of summative assessment on learning. *Assessment & Evaluation in Higher Education*. 35(3), 335–345.

Kachru, B. (Ed.). (1992). *The other tongue: English across cultures* (2nd ed.). Urbana: University of Illinois Press.

Kang, O. (2008). Ratings of L2 oral performance in English: Relative impact of rater characteristics and acoustic measures of accentedness. *Spaan Fellow Working Papers in Second or Foreign Language Assessment, 6*, 181.

Kern, R. G. (1994). The role of mental translation in second language reading. *Studies in Second Language Acquisition, 16* (4). 441-61.

Kim, Y. H. (2009a). A G-theory analysis of rater effect in ESL speaking assessment. *Applied Linguistics, 30*(3), 435–440.

Kim, Y. H. (2009b). An investigation into native and non-native teachers' judgments of oral English performance: A mixed methods approach. *Language Testing, 26*(2), 187–217. doi:10.1177/0265532208101010

Larsen-Hall, J. (2009). *A guide to doing statistics in second language research using SPSS*. New York: Routledge.

Lasagabaster, D., & Sierra, J. M. (2002). University students' perception of native and non-native speaker teachers of English. *Language Awareness, 11*(2), 132-142.

Liu, J. (1999). Nonnative-English-speaking professionals in TESOL. *TESOL Quarterly, 33*(1), 85-102.

Liu, J. (2005). Chinese graduate teaching assistants teaching freshman composition to native English speaking students. In E. Llurda, (Ed.), *Non-native language teachers: perceptions, challenges and contributions to the profession* (pp. 129-155). New York, NY: Springer Science+Business Media, Inc.

Llurda, E. (2005). Looking at the perceptions, challenges, and contributions…or the importance of being a non-native teacher. In E. Llurda, (Ed.), *Non-native language teachers:*

Perceptions, challenges and contributions to the profession (pp.1-9). New York, NY: Springer Science+Business Media, Inc.

Ludwig, J. (1982). Native-speaker judgments of second-language learners' efforts at communication: A review. *The Modern Language Journal, 66*(3), 274–283.

Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing, 12*, 54–71.

Malik, S. K., & Gulnaz, R. (2011). An investigation into students' perceived views about effectiveness of oral presentation at university level. *Language in India, 11*(10), 343-356.

McBrien, J. L. & Brandt, R. S. (1997). *The language of learning: A guide to education terms.* Alexandria, VA: Association for Supervision and Curriculum Development.

McKay, S. (1992). *Teaching English overseas: An introduction*. Oxford: Oxford University Press.

McNeil, A. (2005). Non-native speaker teachers and awareness of lexical difficulty in pedagogical texts. In E. Llurda, (Ed.), *Non-native language teachers: perceptions, challenges and contributions to the profession (pp. 107-128).* New York, NY: Springer Science+Business Media, Inc.

Medgyes, P. (1992). Native or non-native: Who's worth more? *ELT Journal, 46*(4), 340-349.

Medgyes, P. (1994). *The non-native teacher*. London: Macmillan Publishers.

Medgyes, P. (2001). When the teacher is a non-native speaker. In M. Celce-Murcia, (Ed.), *Teaching English as a Second or Foreign Language* (pp. 429-442). MA, Boston: Heinle & Heinle Publisher.

Minnesota State University Mankato. (2015a). *International students* [Website]. Retrieved from: http://www.mnsu.edu/admissions/international/

Minnesota State University Mankato. (2015b). *TESL faculty & staff* [Website]. Retrieved from:

http://english.mnsu.edu/tesl/teslfaculty.html

Modiano, M. (2005). Cultural studies, foreign language teaching and learning practices, and the

NNS practitioner. In E. Llurda, (Ed.), *Non-native language teachers: perceptions,*

challenges and contributions to the profession (pp. 25-43). New York, NY: Springer

Science+Business Media, Inc.

North, B. (1995). The Development of a common framework scale of descriptions of language

proficiency based on a theory of measurement. *System, 23*(4), 445-465.

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing,*

19, 169–92.

O'Loughlin, K. (2001). *The equivalence of direct and semi-direct speaking tests.* Cambridge:

Cambridge University Press.

O'Sullivan, B. (2000). Learner acquaintanceship and oral proficiency test pair-task performance.

System, 28(3), 373-386.

O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance.

Language Testing, 19(3), 277-295.

O'Sullivan, B. (2008). *Modelling performance in oral language tests.* Frankfurt: Peter Lang.

O'Sullivan, B. (2012). Assessing speaking. In C. Coombe, P. Davidson, B. O'Sullivan, & S.

Stoynoff, (Eds.), *The Cambridge guide to second language assessment* (pp. 234-246).

Cambridge: Cambridge University Press.

O'Sullivan, B. (2014). Assessing speaking. In A. J. Kunnan (Ed.), *The companion to language*

assessment (pp. 1-16). Oxford: Wiley.

O'Sullivan, B., & Rignall, M. (2007). Assessing the value of bias analysis feedback to raters for the IELTS writing module. In M. Milanovic, & C. J. Weir, (Eds.*), IELTS research in writing and speaking* (pp.446-476). Cambridge: Cambridge University Press / Cambridge ESOL.

O'Sullivan, B., & Yang, L. (2006). An empirical study on examiner deviation from the set interlocutor frame in the IELTS speaking paper. *IELTS Research Reports, 6*, 91–118.

Paikeday, T. (1985). *The native speaker is dead!* Toronto, Canada: Paikeday.

Prodromou, L. (1992). What culture? Which culture? Cross–cultural factors in language learning. *ELT Journal, 46*(1), 39-50.

Reed, D., & Cohen, A. (2001). Revisiting raters and ratings in oral language assessment. In C. E. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, & K. O'Laughlin (Eds.), *Experimenting with uncertainty: Essays in honour of Alan Davies* (pp. 82–96). Cambridge: Cambridge University Press.

Reves, T., & Medgyes, P. (1994). The non-native English speaking EFL/ESL teacher's self-image: An international survey. *System, 22* (3), 353-67.

Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review/ La Revue canadienne des langues vivantes, 65*(3), 395–412. doi:10.3138/cmlr.65.3.395

Saminy, K., & Brutt-Griffler, J. (1999). To be a native or non-native speaker: Perceptions of 'non-native' students in a graduate TESOL program. In G. Braine, (Ed.), *Non-native educators in English language teaching* (pp. 127-144). Mahwah, NJ: Lawrence Erlbaum.

Santos, T. A. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly, 22*, 69–90.

Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students'

    English writing. *Language Testing, 18*(3), 303–325. doi:10.1177/026553220101800303

TESOL International Association (1999). Non-native English Speakers' Caucus [Report].

    Retrieved from: https://www.tesol.org/read-and-publish/journals/other-serial-

    publications/compleat-links/compleat-links-volume-4-issue-1-(march-2007)/tesol-

    caucuses-are-about-who-we-are

TESOL International Association (2014). *Explore Sustain Renew*. [Annual Report]. Retrieved

    from: https://www.tesol.org/docs/default-source/annual-reports/2014-annual-

    report.pdf?sfvrsn=0

Weigle, S. (1994). Effects of training on raters of ESL compositions. *Language Testing, 11*(2),

    197-223.

Weigle S. (1998). Using FACETS to model rater training effects. *Language Testing, 15*(2), 263-

    267.

Weir, C. J., O'Sullivan, B., & Horai, T. (2004). Exploring difficulty in speaking tasks: An intra-

    task perspective. *IELTS Research Reports, 6*, 1–42.

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. doi:10.1177/0265532209360671

Zitouni, N. (2013). *The use of students' oral presentation in enhancing speaking skill in the

    English language classrooms*. Algeria: Biskra University.

Appendices

Appendix A: Sample Email to Participants

Email to be sent to potential participants:

Dear name,

Thank you for taking the time to review this email. My name is Rubaiyat Jabeen and I am a graduate student in the MA Teaching English as Second Language (TESL) program at Minnesota State University, Mankato. You are invited to participate in a research study that compares native and non-native English speaking instructors' assessment of English as a second language (ESL) students' oral presentations. Note that your participation is voluntary. If you choose to take part in this research, rest assured that your responses will be anonymous. By clicking on the online survey tool below, you will be asked to watch a 12-minute video recording of an ESL student's oral presentation. After watching the video, you will be asked to evaluate the student's oral presentation skills by filling out a rating scale which will take 2-3 minutes. Following this you will then be asked to fill out a brief questionnaire of about 5 minutes related to your views about assessing ESL students' oral presentations.

Qualtrics link: https://mnsu.co1.qualtrics.com/SE/?SID=SV_7WzYWh2BfGq1wX3

Thanking you in advance for your consideration to participate and time you devote to completing the online survey.

Sincerely,

Rubaiyat Jabeen

Appendix B: Online Survey Consent Form

ONLINE/ANONYMOUS SURVEY CONSENT

You are kindly requested to participate in a research study on the assessment of English as a second language (ESL) student's oral presentations. The purpose of the study is to investigate the differences in the assessment practices of native English-speaking (NS) and non-native English-speaking (NNS) instructors in regards to reliability and evaluation criteria when assessing English as a second language (ESL) student's oral presentations.

This study is supervised by Dr. Glen Poupore and conducted by Rubaiyat Jabeen, an MA TESL graduate student in the Department of English at Minnesota State University, Mankato. You were selected as a potential participant in the study because you are an instructor of oral communication or an instructor of ESL students. You will be asked to watch a 12-minute video recording of an ESL student's oral presentation. After watching the video, you will be asked to evaluate the student's oral presentation skills by filling out a rating scale which will take 2-3 minutes. Following this you will then be asked to fill out a brief questionnaire of about 5 minutes related to your views about assessing ESL students' oral presentations. If you have any questions about the research, please contact Dr. Glen Poupore at glen.poupore@mnsu.edu or Rubaiyat Jabeen at rubaiyat.jabeen@mnsu.edu.

Participation is voluntary. You can choose not to participate in this research, and you may stop taking the survey at any time by closing your web browser. Participation or non-participation will involve no penalty or loss of benefits and will not impact your relationship with Minnesota State University, Mankato, or with the investigators. If you have questions about the treatment of human participants' rights and Minnesota State University, Mankato, please contact the Institutional Review Board (IRB) Administrator, Dr. Barry Ries, at 507-389-2321 or barry.ries@mnsu.edu.

Your responses to the survey will be anonymous, and the records of this research study will be kept confidential. The results of the survey will be kept on a secured laptop. It is also assured that any publications and presentations of the results will not include demographic descriptions of individual participants that are detailed enough to make identification possible. However, as the data collection will use online technology, there might always be the risk of compromising privacy, confidentiality, and/or anonymity. For more information about the specific privacy and anonymity risks caused by online surveys, contact the Minnesota State University, Mankato Information and Technology Services Help Desk (507-389-6654) and ask to speak to the Information Security Manager or email servicedesk@mnsu.edu.

The risks you will encounter as a participant in this research are not more than experienced in your everyday life.

There is no direct cost or benefit to you for participation in this research. Participation will cost you only time and you will not receive money to participate. However, results gathered from the study might provide a better understanding of instructors' assessment practices of ESL students' oral presentations.

PARTICIPANT CONSENT

Submitting the completed survey will indicate your informed consent to participate and your assurance that you are at least 18 years of age.

Please do not hesitate to print a copy of this page for your future reference.

MSU IRBNet ID # for this research: 880089

Date of MSU IRB approval: March 14, 2016

Appendix C: Qualtrics Survey 2016-03-05

This survey consists of 3 sections. Please read each instruction carefully and follow the directions. This is not a test so there is no 'right' or 'wrong' answer and you do not even need to indicate your name. The results of the survey will be used only for research purposes so please give your answers sincerely. Thank you very much for your assistance.

Part 1

Please watch the following video just once. There are two ESL students presenting together in the video. You will only assess and score the South Korean student's oral performance. After watching the video, please score her performance using the rating scale.

[Video located here]

Assessment Criteria	Grade Points				
	5	4	3	2	1
1. Communicates the content and his/her thoughts in a **clear and coherent manner**.					
2. **Uses own words and speaks naturally** (not in a memorized fashion or reading from a text).					
3. **Speaks fluently and smoothly** without too many pauses, fillers (e.g. 'uh'), and/or hesitations					
4. Speaks in a manner which is **not too fast or too slow.**					
5. **Pronunciation of individual words/phrases** is clear**.**					
6. **Uses variance in intonation patterns** (not speaking in a monotone voice).					
7. Uses **correct grammar forms.**					
8. Makes good **eye contact** with the audience when speaking (e.g. ensures wide eye contact with all sides of the audience, not frequently looking at notes, looks at audience rather than projection screen, board, or computer screen).					
9. **Speaks loudly enough** so that all audience members can clearly hear.					
10. Effective use of **body language** (posture and					

	gestures).					

Part 2

In this part of the questionnaire, we would like you to tell us how much you agree or disagree with the following statements in relation to assessing the oral presentations of university international students who have English as their second language.

	Strongly Disagree	Disagree	Neither Agree or Disagree	Agree	Strongly Disagree
I give (or would give) a lot of importance to correct pronunciation in comparison to other evaluation criteria.					
I give (or would give) a lot of importance to correct grammar in comparison to other evaluation criteria.					
I give (or would give) a lot of importance to fluency (smooth flowing speech without too many pauses, hesitations, and/or filler words such as 'uh') in comparison to other evaluation criteria.					
I give (or would give) a lot of importance to eye contact in comparison to other evaluation criteria.					
I give (or would give) a lot of importance to body language (posture and use of gestures)					

in comparison to other evaluation criteria.					
I give (or would give) a lot of importance to clarity of content delivery in comparison to other evaluation criteria.					
I think oral presentations are a good assessment tool to measure the speaking skills of international students who have English as their second language.					

Part 3

Please answer the following biographic questions.

What is your gender?

Male _____

Female _____

What is your age? _____

What is your nationality? _____

What is your first language? _____

Have you assessed the spoken English of an international student before?

Yes _____

No _____