RATERS' DIFFERENT JUDGMENT OF INTELLIGIBILITY IN THE ENGLISH
PLACEMENT TEST AT UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN

BY

JI EUN LEE

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Arts in Teaching of English as a Second Language
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Adviser:

Professor Fred G. Davidson

ABSTRACT

This study investigated raters' judgments of test-takers' intelligibility on Phase I of the EPT (English Placement Test), and how their various teaching experiences and different native languages affected their judgments. The EPT is administered to incoming international students whose TOEFL scores do not meet campus or departmental cutoff scores at the University of Illinois at Urbana Champaign. Phase I of the EPT oral section is designed to measure students' intelligibility, and employs a group discussion format. Nine raters with different teaching experience and native languages were interviewed, and their responses were analyzed qualitatively.

Results show that all the raters interpreted oral intelligibility differentially, and raters' different understanding of the constructs underlying intelligibility resulted in different rating scales. This result wasn't affected by raters' teaching experience or language backgrounds. Rather than these, familiarity with a specific accent or rater's individual ideas about intelligibility had more influence on their judgment. When they were asked about which specific features contribute more to intelligibility, most raters responded that segmental factors and lexical stress were important factors for their judgment. This might be because English Placement Test is a tool for assessing the 'word level' intelligibility.

Phase I of the English Placement Test uses a one-on-one interview format or a group-administered format at raters' discretion. Most raters reported that they prefer the one-on-one interview format. They argued that even though the group-administered format had the benefit of reducing time and expense, it would threaten the validity of the

test because of the possibility of copying another test taker's response or of getting nervous due to the situation in which test takers had to speak in front of group members.

Providing a clarified and unified definition of intelligibility in the recalibration session is suggested. Additionally, raters still need to have sufficient knowledge of what the group oral discussion format is and know how to manage this format.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

I. INTRODUCTION

This chapter provides a brief introduction to the study, along with the research questions it aims to investigate.

1.1 Context

The concept of intelligibility has been placed as central to much of the research that has examined pronunciation teaching and testing, and has more recently emerged from the study of World Englishes. However, there is neither an agreed definition of intelligibility among researchers nor a consensus on how to measure it. For example, some researchers make a relatively clear distinction between "intelligibility" and "comprehensibility" while others use these two terms interchangeably. Bansal's (1969) definition of "intelligibility" is based on phonetic and phonological criteria. He argues that to be intelligible, the speaker must articulate his sounds and words clearly, minimizing the hearer's efforts to understand what word is meant (Bansal, 1969: 15). Smith and Rafiqzad (1979) develop the distinction between "intelligibility" and "comprehensibility," defining "intelligibility" and "comprehensibility" with different sets of parameters. In their large-scale study intelligibility is measured by the extent to which someone is able to represent words heard, while intelligibility is regarded as including lexical choice and grammatical structures. On the other hand, other researchers use of "intelligibility" is not compatible with that of Bansal (1969) or Smith & Rafiqzad (1979).

Bamgbose acknowledges "intelligibility" as a more complex matter than just phonological representation of words. He states that intelligibility is comprised of various factors "recognizing an expression, knowing its meaning, and knowing what that meaning signifies in the sociocultural context."(Bambose, 1998, p. 11). James is one of the researchers who uses "intelligibility" with a different meaning. He refers to "intelligibility" as "the accessibility to the basic, literal meaning, and the propositional content encoded in an utterance" (James, 1998, p.212).

Due to the various uses of the term "intelligibility," it seems logical that there is no universally accepted way of measuring it (Munro & Derwing, 1999). What is important here is that inferences made about test-takers' abilities based on their test scores depend on the extent to which the scores are determined by the construct which one desires to measure (Ockey, 2009, P.2).

When intelligibility is defined as word and utterance recognition, the segmental phonological factors that affect intelligibility have been discussed to some extent in many studies (Gimson, 1970; Jenkins, 2007; Munro & Derwing, 1995; Schairer, 1992). Researchers, however, do not agree on which specific aspects of pronunciation are most crucial for intelligibility. Several studies attempt to distinguish important phonological features from relatively unimportant ones for acquiring intelligibility. Research results have reached no firm conclusion due to differences of methodology (Munro and Derwing, 1995). Results also vary according to the context of the studies and whether they employ native speaker listeners or include non-native speaker interaction.

Additionally, test task is a factor that might affect the measurement of intelligibility. With the recently growing use of performance-based speaking tests, intelligibility is

assessed in the context of various speaking tasks. When we incorporate performance-based speaking tests in assessing intelligibility, other factors rather than test-takers' intelligibility could influence the assignment of test scores. Other factors might include test-takers' underlying competence, raters, scale criteria, speech samples, task qualities, interlocutors and the speaking performance. When intelligibility is the intended construct, the other factors could be considered as construct-irrelevant variance.

Raters' grading could also be affected by various factors. Their experiences, language backgrounds, and even their individual characters could all affect results. Task type is also considered to have consequences on scores since it provides a different context for test-takers' performance. Traditionally, in second language speaking tests one-on-one interviews were conducted to measure test-takers' oral abilities. However, more recently, due to the limited resources of time and expense, alternative ways to assess test takers' speaking abilities have appeared. The group oral test format is one of these alternatives because it reduces the time and expense of test administration. However, construct irrelevant variables that could threaten the validity of the tests because of a group format context have caused concern. For example, the personal characteristics of test takers' group members have been reported as a threatening variable.

1.2 The Study

The purpose of this study is to explore the two-phase structure of the oral test section of the English Placement Test (EPT) administered by the University of Illinois at Urbana-Champaign (UIUC). The EPT is administered to incoming international students whose

TOEFL scores do not meet campus or departmental cutoff scores. The primary purpose of the EPT at UIUC is to provide a basis to require students to take, recommend that they take, or exempt them from taking appropriate ESL writing and/or pronunciation courses. Based on university policy, new international students are evaluated (by a campus admissions officer) for evidence of English language ability and required (or not) to take the EPT. Hence, there is a two-tier requirement: first, is the EPT required? If so, does its result require any ESL course(s)? Failure to fulfill either requirement can impact a student's ability to register for courses. Therefore, the results of the placement test are of great importance, and the stakes seem relatively high.

The EPT consists of two parts: a written test and an oral interview. The writing test requires students to write an academic essay based on a reading passage and a related lecture. After writing an essay, the students are asked to discuss a related topic with peers. During this discussion, an oral screening measure is administered, which is known as Phase I of the oral test. At the Phase I oral screening, students are given a topic on which to speak for three minutes. If students speak intelligibly, they will be exempted from further oral testing. Otherwise, students will be required to take Phase II of the oral test scheduled in the afternoon on the same day as written test. However, it is not always the case that Phase I and Phase II are not administered separately. When the resources permit, test takers take Phase I and Phase II together. In this case, test takers are asked general questions at the beginning of the oral test. In this study, only Phase I of the oral test will be investigated.

More precisely, Phase I is an oral screening measure which is designed to sort out the test takers into three groups through a short oral interview. When test takers are

considered "intelligible," they are exempted from taking any ESL pronunciation courses. If raters are unable to make a decision about a test-taker's need to take pronunciation classes in a short conversation, the test taker is categorized as "conditionally intelligible" and asked to take the full pronunciation test, Phase II. If test takers are judged as "unintelligible," they are required to take a pronunciation classes. Phase II begins with a full one-on-one oral interview, providing the test taker another chance to get exempted from the requirement to take pronunciation classes. If raters still are unable to make a solid exemption decision through the oral interview, the test taker goes on to sentence/phrase/dialog reading. According to the results, he or she is recommended to take an ESL pronunciation class or required to do so.

The first phase of the oral EPT adopts intelligibility as its central measure (Issac, 2008). This test includes a subjective, word-based definition of intelligibility to assess non-native speaking students. The primary purpose of the oral section of the EPT is to determine if the students need to take an ESL pronunciation course and to place them into an appropriate ESL pronunciation course. (see Figure 1)
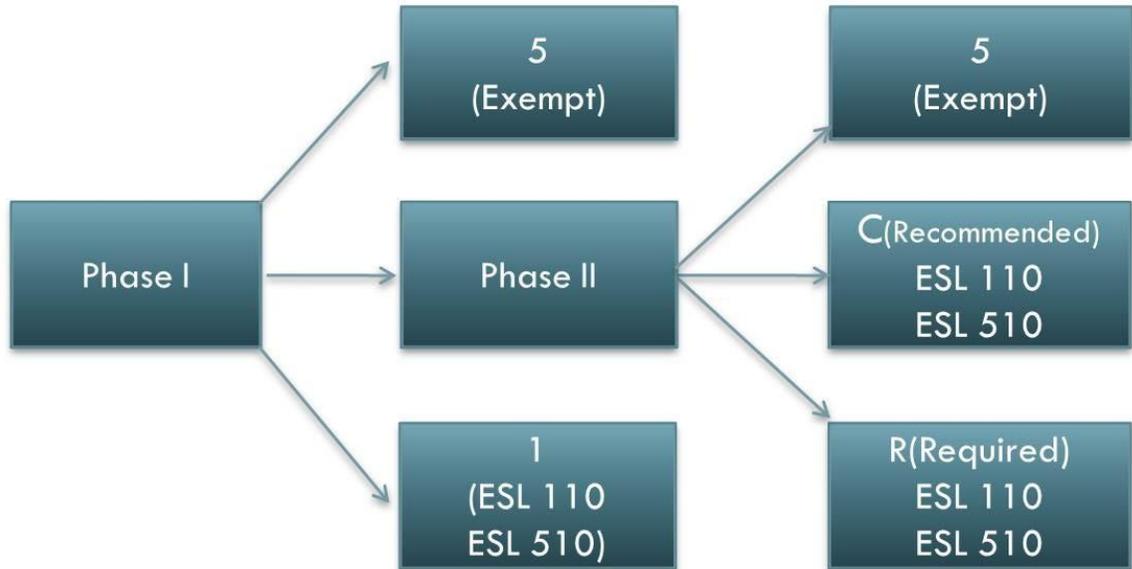
Figure 1. The structure of the EPT oral section when operated as a two-phase exam

For several years in the Phase I of EPT oral section there have been anecdotal reports of inconsistency among raters in assessing intelligibility. For example, one of the fall semester, 2009 raters, an English native speaker with a couple of years of experience teaching ESL pronunciation courses, reported that he assigned significantly lower scores than other raters who were non-native speakers of English or who had relatively less experience in assessing and teaching ESL learners. His grading appeared to be distinctively low compared to other native speaker graders with many years of grading experience.

Rater variability has been widely discussed in the related research, with a number of studies focusing on the effects of raters' diverse linguistic and professional backgrounds (Brown, 1995; Chalhoub-Deville, 1995; Chalhoub-Deville & Wigglesworth, 2005; Fayer & Krasinski, 1987; Hadden, 1991). Generally, raters with teaching

experience and language backgrounds other than English were shown to be harsher in their assessment than non-teachers and native speakers (Fayer & Krasinski, 1987; Hadden, 1991). However, a couple of studies showed contradictory results (Brown, 1995; Chalhoub-Deville, 1995). Most studies on rater variability in second language oral assessment have dealt with overall oral ability such as fluency, accuracy, communicability and grammar as well as pronunciation. Therefore, it seems difficult to compare these results with this study, which focuses on only assessment in pronunciation and intelligibility.

Besides rater variability, there are other possible triggers for rater inconsistency. Different understanding of the construct of intelligibility could cause raters to behave differently in their assessment. That is, raters with diverse backgrounds might tap different constructs of underlying intelligibility, possibly resulting in a different interpretation of rating scales. In addition, raters might have contradictory, or at least not identical, perspectives about which phonological features contribute more to intelligibility. Furthermore, as previous studies reveal, familiarity with a specific accent or having the same language background as test-takers would affect raters' assessment of their performances. Among these possible explanations, what are the most logical arguments that would explain the rater inconsistency in this current research?

The purpose of this study, therefore, is to explore how raters with diverse backgrounds behave differently in assessing test-takers' intelligibility. To do this I will examine how each one defines intelligibility, understands the construct of intelligibility and applies the rating criteria. I propose to determine which phonological factors of test-takers' speech contribute to their ratings. And considering that not all raters used the same

format in the Phase I of the EPT oral section, I will record their preference of testing

format, reasons for their preferences, and problems they experience while employing the

group format.

II. Literature Review

2.1. Intelligibility

The notion of intelligibility has been considered an appropriate goal for the teaching of pronunciation for several decades. As far back as 1949 Abecrombie argued that "language learners need no more than a comfortably intelligible pronunciation" (p.120). In this argument, he defines 'comfortably intelligible' as 'a pronunciation which can be understood with little or no conscious effort on the part of listener' (p.120). Several researchers advocate Abecrombie's (1949) view and adopted intelligibility as a plausible aim for second language pronunciation teaching.

Since then there has been no agreement among researchers on how to define intelligibility. Kenworthy (1987) explains 'comfortably intelligible' as being understood by a listener at a given time in a given situation (p.13). In her definition of being 'comfortably intelligible,' the more words the listener is able to accurately identify, the more intelligible the speaker is. Kenworthy claims that in order to be comfortably intelligible, the listener is able to understand the speaker without the need to ask for repetition or clarification.

Morley (1994) examines intelligibility with more detail, and adds the terms 'functional intelligibility' and 'overall intelligibility' to 'comfortably intelligible' and 'intelligibility.' She maintains that if non-native speakers are not "comfortably intelligible" they often are not able to maintain spoken interaction. While she does not specifically define 'comfortable intelligibility' and 'intelligibility,' functional

intelligibility is taken to mean how easy it is to understand a non-native's utterances. Both Morley (1994) and Kenworthy (1987) equate intelligibility with understandability to some extent. Specifically, Morley (1994) relates intelligibility to the broader goal of communication (Isaacs, 2008). Her Speech Intelligibility/communicability Index (1994) shows that she assesses intelligibility in terms of accent and its effect on listener perception.

Whereas Morley relates intelligibility to communication in terms of accent and its effect on listener perception, a number of studies on intelligibility in English have paid most attention to pronunciation. This line of studies makes a clear distinction between intelligibility and comprehensibility. That is to say, the research does not consider communicative success when examining intelligibility. Smith and Rafiqzad (1979) define intelligibility as accessibility of pronunciation. They measure intelligibility as the extent to which speakers are able to represent words. Smith (1980) develops the distinction between intelligibility and comprehensibility, and regards comprehensibility as involving a different set of parameters from intelligibility. Comprehensibility includes how to use lexical items and organize grammatical structures. He adds one more notion, interpretability, which is critical for successful communication. Interpretability is defined as listeners' ability to discern the motivation or purpose of utterances (Nelson, 2008). This notion is closely related to pragmatics.

Jenkins (2000) adapts Smith and Nelson's (1985) definition of intelligibility. She uses this term as the ability to produce and receive phonological form, and regards intelligibility as a prerequisite for successful comprehensibility or interpretability. Bamgbose (1998) also uses Smith's (1980) terminology, but his definition does not

correspond precisely. On the other hand, James (1998) employs intelligibility for semantic criteria. He refers to intelligibility as the accessibility of the basic, literal meaning and propositional content encoded in an utterance (cited by Jenkins, 2000:70).

Munro, Derwing, and Morton (2006) use the same ideas of intelligibility and comprehensibility as Smith does, but add the concept of "accented-ness" as well. According to Nelson's (2008) interpretation, intelligibility is defined as the actual understand-ability of speech production, while comprehensibility is interpreted as listeners' self-perception of difficulty in processing what they hear. In their study accented-ness is related to how the patterns of oral production are different from what listeners are accustomed to.

Derwing and Munro (1995, 1997, 2006) distinguish intelligibility from accented-ness and comprehensibility. They define intelligibility as the extent to which a speaker's utterance is actually understood, whereas comprehensibility is the listeners' estimation of difficulty in understanding an utterance. Accented-ness is the degree to which the pronunciation of an utterance sounds different from an expected production pattern. In order to explore the relationships among intelligibility, perceived comprehensibility, and accented-ness, Munro and Derwing (1995) compared transcription scores for intelligibility with accent and perceived intelligibility ratings. They focused on various features such as phonetic, phonemic, and grammatical errors, and quality of intonation in Mandarin speakers' English production. The results show that for most listeners, accent ratings were significantly correlated with all error types and intonation ratings. Comprehensibility ratings correlated with quality of intonation for the majority of listeners, but phonetic errors were correlated less significantly. Intelligibility was not

correlated with phonetic errors, although it was correlated with phonemic and grammatical errors and quality of intonation for only a small number of listeners. The researchers argue that although intelligibility, perceived comprehensibility, and accented-ness are interrelated, they are considered as separated dimensions. Their study was extended by examining listeners' reactions to speech from a variety of language backgrounds (1997). They analyzed accent, comprehensibility ratings, and transcriptions of accented speech from Cantonese, Japanese, Polish, and Spanish ESL intermediate students. Familiarity with various language backgrounds and native speakers' ability to identify the particular language backgrounds were also examined. In addition, the researchers compared the relationship observed in the speech of high-proficiency speakers with the speech of intermediate non-native speakers. In general, accent was rated more harshly than perceived comprehensibility, which in turn was rated more harshly than actual intelligibility scores. Results of correlational analyses indicate that the three dimensions are clearly related but are not equivalent. Compared to their own previous study (1995), there are fewer significant correlations between accent ratings and the accent features of phonemic errors, grammatical errors, and native-ness of prosody. These errors had a limited effect on comprehensibility.

Derwing and Munro replicated and extended their own work (1997) to examine the extent to which listeners from a variety of L1 backgrounds share a response to L2 speech. They compared responses from listeners familiar with the pronunciation patterns of the speakers with those who have no such familiarity. Cantonese, Japanese, Mandarin learners of English and native speakers of English participated in transcription and ratings of the same utterances of 4 different L1 background speakers, which had been used in

their previous study (1997). The results indicate that even though a variety of factors including properties of the speech itself, the L1 backgrounds of the listeners, or bias against accents might have been influential, different listener groups showed similar results in their comprehension and evaluation of nonnative utterances. Listeners from three different L1 backgrounds showed moderate to moderately high correlations in their responses on intelligibility, comprehensibility, and accented-ness. The effect of familiarity on these three dimensions was small. Nonnative listeners did not find speech produced with their own accent to be more intelligible, more comprehensible, or less accented than speech produced with less familiar accents. Finally, intelligibility, comprehensibility, and accented-ness scores did not vary to a great extent between native English and L2 listeners. The researchers concluded that the effects of L1 background and experience with a particular type of accent did not contribute greatly to the ability to understand L2 speech.

In summary, it seems that there has hardly been consensus among researchers how they define intelligibility, in particular about whether the definition of intelligibility include being able to understand speaker's utterance or intended message, and how much familiarity to a specific accent determine the intelligibility.

The EPT oral section adopts the notion of intelligibility under ~~the~~ a specific circumstance. The raters are trained on specific phonological features to answer the following question for Phase I of the EPT: "Can I understand everything the speaker says?". More importantly, if the raters are in doubt, they are supposed to send test takers to Phase II. Hence, in this study, it is likely that a new operational definition is needed. The operational definition will be that the rater must have full, immediate understanding,

and if the rater is in doubt, then a further assessment will happen in the Phase II. This new operational definitions will show how the EPT is substantially different from definitions of intelligibility from the previous research in the language testing literature where the term "intelligibility" is operationalized in a single-stage testing event.

2.1.1 Contributory Factors

In Phase I of the EPT oral section, raters judge test-takers' pronunciation based on their overall impression, while in Phase II they attempt to count the various types of errors that occur. The criteria used for Phase I are based on intelligibility, and raters are instructed to judge errors mainly as they affect intelligibility. Since these are holistic judgments, the question that arises is whether native speaker judges react equally to deviance in all major areas of pronunciation, or whether each area carries a different weight in influencing the scores that are assigned. The major areas of pronunciation are segmentals, suprasegmentals, and syllable structures. Deviance in segmentals involves errors in consonants and vowels, such as the substitution of one sound for another or the modification of a sound. Errors in suprasegmentals include deviations in patterns of stress and intonation which affect phrasing and rhythm. Syllable structure errors involve the addition of a segment or syllable, the deletion of a segment or syllable, or the reordering of segments in syllables, the most common types of errors being consonant deletion and vowel insertion (Anderson-Hsieh, Johnson & Koehler, 1992). However, there has been

no agreement as to which specific aspects of pronunciation contribute most to intelligibility.

Several studies attempt to establish hierarchies of pronunciation errors (Gimson, 1970; Jenkins, 2007; Munro & Derwing, 1995; Schairer, 1992). Regarding the relative importance of segmental factors, Gimson (1970) maintains that accurate production of consonants is more helpful for listeners' comprehensibility, while Schairer (1992) argues that native-like production of vowels is more essential for intelligibility. Jenkins (2007) maintained that some phonological features are more important than others for successful communication particularly among non-native speakers of English with different L1 backgrounds.

Most research has been concerned with the relative importance of prosody versus segmentals. Some research evidence suggests that suprasegmentals play a more important role in intelligibility than segmentals. Anderson-Hsieh, Johnson, and Koefler (1992) compared the relative contribution to intelligibility between prosody, segmentals and syllable structure. Their analysis of 60 reading passage speech samples from 11 different language groups indicated that while deviance in segmentals, prosody, and syllable structure all showed a significant effect on the pronunciation ratings, the score for prosody was most significantly related to the overall score for pronunciation. In a related finding, Derwing, Munro, and Wiebe (1998) found that suprasegmental instruction had a greater effect on second language learners' communicative performance than segmental instruction. James (1976) examined the relative effect of articulation and prosody on native speakers' perception of accented-ness in the speech of English speakers learning

French. The ratings of the native French listeners showed that good intonation and poor articulation was more acceptable than was poor intonation with good articulation.

On the other hand, some studies have reported results contradictory to these findings. Fayer and Krasinski (1987) investigated the reactions of native speakers of English and native speakers of Puerto Rican Spanish to the speech of Puerto Rican ESL learners. The listeners were asked to judge the intelligibility and irritability of the several aspects of Puerto Rican ESL learners' oral communication including grammar, intonation, pronunciation, hesitation, and lexical errors. The researchers found that deviant pronunciation and hesitation were more distracting than incorrect intonation. In general, however, more evidence has supported the primacy of prosody over segmentals, the mixed results seem to be due to the difference of target language and the methodology adopted.

Besides research into the comparison of segmental and suprasegmental factors, Jenkins (2007) conducted empirical research on the causes of intelligibility problems in non-native speaker interactions and was able to identify some core phonological features critical for successful communication among non-native speakers. These features included all consonant sounds except for the dental fricatives and dark /l/, vowel length contrasts, initial and medial consonant clusters, and tonic stress. In contrast to these features, word stress was found to be unimportant for communication between non-native speakers.

In addition, a number of studies have dealt with prosody or suprasegmental factors as a unitary construct. However, prosody has various constituents such as lexical stress, intonation, and the relative duration of weak and strong syllables. It is possible that these

various variables have different effects on intelligibility, but lexical stress was reported as playing a central role in processing words. Cutler and Clifton (1984) in their L1 study showed evidence that certain types of stress misplacement seriously impair intelligibility. In particular, intelligibility was affected when the stressed syllable was shifted rightward.

As reviewed so far, the relationship between these different aspects of pronunciation seem complex, how raters integrate the judgment of these aspects and reach a final decision about test takers' pronunciation ability depends on the individual context, including the purpose of the test.

2.2. The Group Oral Format

Most of the performance-based speaking test research has dealt with the one-on-one interview or simulated tape-mediated interviews. Compared to interest in traditional oral test formats, the group oral test format has received relatively little empirical attention from language testing researchers, although this is starting to change.

According to Ockey (2009), previous research has generally defined the group oral test as a speaking test format "in which at least three second language learners discuss a topic while raters sit outside the group and give scores to the individual test-takers without participating in the discourse" (p. 162, Ockey, 2009). This format has been considered attractive in the second language speaking literature due to its cost-effectiveness for large-scale oral proficiency testing. It has been reported to be more time-efficient and cost-effective than traditional interview tests (Bonk & Ockey, 2003). The group oral test format enables raters to assess a relatively large number of students at

one time, which seems an economical way to assign speaking scores. It reduces raters'

burden since they don't need to lead a conversation by asking questions or controlling the

discourse, so they are able to focus on test-takers' performance. Peer discussions help

raters to be more consistent in their rating compared to other formats like the oral

interviews where raters play a role in the test discourse (Folland and Robertson, 1976;

Ockey, 2001). The group oral discussion format has also been reported to have positive

washback to the classroom in contexts where more communicative language learning

goals are needed (Shohamy et al., 1986). In these contexts, the group oral discussion

format is connected to classroom practices since many learning activities take the form of

small group collaboration and discussion. Thus, this testing format seems to reflect

classroom environments and arguably, the real world.

The group oral test as a second language speaking test format has been discussed

in the literature for several decades. Wilkinson (1968) was one of the first to suggest its

use. However, in this early research, there was no clear evidence to validate score-based

inferences. Folland and Robertson (1976) also suggested the group discussion as one of a

battery of oral tests that could elicit sizable evidence of oral ability. They maintained that

this format had the advantage of reducing the examiner's burden since they didn't have to

lead the conversation. Since then, several studies have investigated the various facets of

the group oral test format and attempted to determine the extent to which these facets

contribute to test scores in order to validate the score-based inferences yielded from the

test.

Fulcher (1996) compared a group discussion test with one interview testbased on

the description of a picture followed by discussion, and another discussion of a reading

text using questionnaire techniques and retrospective reports from students. In terms of artificiality of the tests and perceptions of validity, responses from the students indicated that engaging in group discussion enabled them to speak with more confidence and say what they wanted more easily rather than responding to an examiner in an oral interview. However, this result was not supported by discourse analysis. The group discussion was also considered as reducing the test-takers' anxiety during the test. More importantly, Fulcher made arguments about the generalizability of test scores among different tasks. Both a G-study and a Rasch partial credit analysis showed that other sources of variance (such as task and rater) had no more significant effect on scores than students' speaking ability, and all three tasks (two interview tests and one group discussion test) were operating on a unidimensional scale (P 36, Fulcher, 1996).

Bonk and Ockey (2003) agree that the group oral test format might be appropriate for certain test situations or as part of a battery of oral tests. They investigated the extent to which the test-taker, rater, prompt, and rating scale contribute to score variance in the group oral. The researchers conducted two separate group oral tests. The whole process of the tests included watching a video in the test-takers' first language (Japanese), which explained the test procedure, and then discussing a topic for approximately ten minutes. Test groups consisted of 3 or 4 members, and two raters rated them individually for pronunciation, fluency, grammar, vocabulary, and communication strategies. Through a many-facet Rasch analysis, the researchers claimed that examinee ability was the most substantial facet, followed by rater severity, and then task difficulty. Even though rater differences in severity were generally large, they were not stable over time for individuals. The researchers reported that test-takers could be reliably separated into 2-3

proficiency levels through group oral tests. Their study suggests that the group oral could be a valid task for measuring oral abilities.

Bonk and Van Moere (2004) investigated the effects of prompt, rater, gender of test-takers, proficiency levels and personal characteristics of candidates on scores in the group oral test. 1055 test-takers in groups of three or four at a Japanese university were rated by two trained raters. Five areas of oral ability were assessed: pronunciation, fluency, vocabulary, and communicative effectiveness. Proficiency level of test-takers was based on teachers' evaluation, and students' levels of shyness were measured by ten Likert scale items that were completed by each test-taker after the tests. Study results indicated that gender and mean proficiency level of a group were not significant predictors of scores. Rater effect was relatively large, but prompt had only a small effect on scores. Van Moere (2006) examined the response validity of the test and consistency across tasks, occasions, and raters in order to test the overall validity of the group oral test. He investigated a group oral test as one part of an in-house English proficiency test for placing students, evaluating students' progress, and providing information for the development of the English language curriculum. The results indicated that candidates gave positive reactions to the group oral test format, and differences in rater severity did not contribute significantly to candidates' test score variation. The G-study revealed that differential rater severity contributes little to a candidate's test score variation. Instead, variations in examinee performance were found to be more responsible for different test scores. Topic and prompt were not significant factors.

Although several studies have found some evidence to validate score-based inferences yielded from a group oral test, the question remains whether this test format

allows candidates to show the full range of their linguistic knowledge. Thus, the validity of score interpretation has still been questioned (Ockey, 2009). One such study conducted by Shohamy et al. (1986) correlated a group oral test with other traditional interview tests in order to examine the validity and reliability of the group oral. They compared the group oral test to three different one-on-one oral tasks, namely a discussion, a reporting task, and a role play, and found that the group oral tests had the lowest correlation with the other test scores. This result seems to show that the group oral test measures a different kind of language ability than the one-on-one tasks. But what construct each test assessed was not clearly defined, and the extent to which the variance was related to the construct was not shown in this study.

Hildson (1995) also cast doubt on the validity of the group format when she investigated whether the group oral test measured the six constructs included in the curriculum by transcribing the two group oral tasks. Transcript analysis found that only one category (imparting and seeking factual information) was assessed though this test format.

Other researchers have questioned the reliability of the group oral test. Nevo and Shohamy (1984) compared the group oral test format to the oral interview, a role-play test, and a reporting test. In their study, language testing experts reported that the group oral test ranked highest in the category of utility standards, but the test was scored as lowest in accuracy. The group oral test placed in the middle with respect to feasibility and fairness. The experts seemed to consider the reliability of this test as suspicious, but their reasons were not made clear in this study. In more recent research, the validity of the group oral was questioned again.

He and Dai (2006) analyzed the discourse of 60 Chinese university students in a group oral test. They reported that very few communication strategies such as negotiation of meaning, checking for comprehension, asking each other for opinions were used. In addition, the test-takers were shown to think of the graders as their target audience, thus they talked to examiners rather than the other candidates. This result revealed that the test-takers did not have sufficient understanding of the group oral format in which they were to perform, and they complained that they did not have enough time to demonstrate their oral proficiency. Therefore, the researchers doubted the validity of the group oral format.

Such doubts are the reason that language testers are intuitively concerned about measuring students' oral ability in interactive groups. They understand that the great number of uncontrollable variables such as gender, age, status, friendship, shyness, talkativeness, opportunities for taking the floor, willingness of individuals to participate, and different proficiency levels in the group can all be factors. Moreover, they worry about unknown effects related to these variables. (Van Moere, 2006, P. 414). Influence from interlocutors has also been observed in the oral interview format, but this construct irrelevant variable could be controlled or minimized by rater training.

"Personality" is another variable in group testing that researchers have investigated. Bonk and Van Moere (2004) conducted a study to examine the effects of test-takers' shyness on their scores in a group oral test. Their study involved 1055 students who were put in 322 groups of three or four. They used psychological questionnaires, which were administered immediately after the tests. The results showed that shyness was a small but significant predictor of scores on the test, and the shy test-

takers got lower scores on the group oral, even though test-takers' proficiency levels were taken into account.

Berry (2004) also addressed the issue of personalities in the group oral test. She used the Eysenck Personality Questionnaire to examine the effects of a group's level of extroversion on a test-taker's score. Candidates read a short research-based text and then discussed the research with the other group members. An analysis of covariance indicates that extroverts received the higher scores when they discussed in a group with a high mean level of extroversion, and introverts got the lower scores when their interlocutors were less extroverted. Her research finding supports the hypothesis that both introverts and extroverts score higher when they are placed in a group whose members are more extroverted.

In a related study Ockey (2009) explored the extent to which test-takers' personal assertiveness affects the scores of others in the group, a possible factor of construct-irrelevant variables. The study conducted MANCOVA analyses to determine the extent to which assertive test-takers' scores are affected by their group's level of assertiveness, and the extent to which non-assertive test-takers' scores are affected by their group's level of assertiveness. The analyses indicate that assertive test-takers got higher scores than expected when grouped with only non-assertive test-takers, and lower scores than expected when grouped with only assertive test-takers. The study found no evidence of an effect for grouping based on assertiveness for non-assertive test-takers' scores.

Another variable targeted by a few studies is talkativeness. For example, Van Moore and Kobayashi (2004) examined how talkativeness contributed to the scores of candidates in 50 group oral discussion tests. They found that there was no overall group

effect related to talking more, but that raters assigned higher scores to candidates who attempted to speak more words or longer turns, regardless of their proficiency level. The various proficiency levels of candidates in a group, however, might be one other important variable. Bonk and Van Moore (2004) included two different proficiency levels in their regression model. They found that neither the groups' mean proficiency level nor the degree of variation in proficiency levels in the group changed the scores that examinees were expected to receive.

As reviewed so far, these studies yielded mixed results about which facets contribute to group oral test scores. It is fair to conclude, however, that there is a rater effect, and that when test-takers do not understand the procedure of the group oral test, the test task itself influences the scores. In addition, various variables such as test-takers' personalities and language proficiency levels have been reported to have effects on the scores, even though these effects were small and the extent to which these variables contributed to the scores was different. Therefore, a validation argument for the group oral test needs more evidence to support inter-rater reliability and to indicate how personal characteristics of test-takers affect their scores.

Many varieties of assessments have been designed to test individual oral ability in groups. Groups of test-takers could be assessed with trained interlocutors participating in the discussion, or they could be tested with other candidates in a group and no intervention of raters. Interlocutors can play a variety of roles in a group test. They may ensure that all the test-takers participate in a group discussion, control the topic, or deploy moves to keep candidates on task (Ockey, 2006).

Phase I of the EPT oral section adopts the group oral test format in part. However, there is a critical difference between this test and the group oral tests that have been investigated by previous research. In this test, test-takers are not allowed to interact with each other in most cases, and they usually interact only with the raters – generally, there is no interaction between test-takers. The raters control the topic, and ensure that every test-taker talks in order to get enough speech samples for measuring test-takers' intelligibility. Phase I of the EPT oral section takes advantage of the time-efficiency and cost-effectiveness of the group oral test format. However, since the test is conducted in a group situation, we can assume that some variables reported in the literature to influence scores would also have an effect on Phase I of the EPT oral section. Test-takers' personalities might have an effect on their performances, and there could be a rater effect. Raters might need training on how to administer the group test and to assign scores , which might not be needed when other tests of oral ability are employed.

Besides the fact that interaction between test takers rarely occurs during Phase I of EPT oral section, there is another important difference between this test and those discussed in the research just reported. The group oral exams I have reviewed here measured overall oral abilities including fluency, grammar and sociolinguistic ability as well as pronunciation. By contrast, Phase I of EPT oral section only measures a specific aspect of oral ability, intelligibility. It is designed to exempt the intelligible test takers and to identify those who are deemed unintelligible. These students are sent on to Phase II of EPT oral section, a full and detailed pronunciation test. To my knowledge, there has been no research examining the oral test format with a two-phase oral test where the first phase is aiming only screening the test takers and the second one is actual placement.

Therefore, the group setting in the EPT Phase I should be distinguished from the research I have reviewed so far, and these research findings are not necessarily comparable or applicable to my research.

III. Methodology

3.1 Participants

A total of nine raters participated in this study. These nine raters were selected

from teaching assistants who had rated test-takers in the EPT oral section. All the raters

have different teaching and rating experiences as well as language backgrounds. Four

raters who had assessed test-takers in the EPT oral section for more than three years were

categorized as experienced raters. They had all taught ESL pronunciation courses for

several years, and they are also native speakers of English. The other five raters were

categorized as less experienced raters. Two of them are non-native speakers of English,

one a native speaker of Korean, and the other a bilingual in Japanese and English. These

two raters had a couple of rating experiences in the EPT oral section, but they had no

experience teaching ESL pronunciation courses. The other three raters participated in the

oral section for the first time in the fall semester of 2010. They are native speakers of

English.

All the raters who participated in this study had taken courses in English as a

second language, English phonology and phonetics, and testing second language learners.

They had also spent many years in teaching ESL students. Three of the experienced raters

are doctoral students in the Department of Linguistics or Educational Psychology, and the

other experienced rater had already received a MATESL degree. Of the five less

experienced raters, four are studying Teaching English as a Second Language in the

department of Linguistics, and the other less experienced rater is currently a doctoral

student in the Department of Educational Psychology, having already received a

MATESL degree.

Table 1. Information about participants

	Native Language	Teaching Experience	Testing Experience (EPT oral section)	Education
1	English	6 years	6 years	Educational Psychology PhD student
2	English	5 years	5 years	Linguistics PhD student
3	English	4 years	4 years	Curriculum & Instruction PhD student
4	English	3 years	3 years	TESL MA student
5	Japanese/English	None	1 semester	East Asian Language and Culture MA student
6	Korean	None	2 years	Educational Psychology PhD student
7	English	None	1 semester	TESL MA student
8	English	None	1 semester	TESL MA student
9	English	One semester	1 semester	TESL MA student

3.2 Research procedure

The framework of this study was based on a qualitative research design. This study employed in-depth interview methods because they would provide a comprehensive and diverse analysis of rating behavior and offer the possibility of examining the process that raters go through in their assessment of test-takers' intelligibility and pronunciation.

The nine raters were interviewed individually from March to September in 2010. These interviews took place a few days after or a month after they had rated oral interviews. The interviews each lasted thirty minutes to one hour. Each interview reviewed the Phase I scoring procedure using the following stages: 1) rating the students' intelligibility according to the four point rating scale; 2) Justifying and explaining those ratings by providing their own evaluation criteria. If the researcher did not understand or wanted further information about the scoring procedure, she contacted the rater individually via e-mail.

The interview questions were not strictly predetermined. According to each rater's answers the questions were slightly different at the researcher's discretion. In general, the questions sought to learn what the raters think the purpose of Phase I is and what rating criteria they rely on during rating. In order to investigate with more depth the features that influence raters to judge test takers' speech differently, questions were asked about raters' own definition of intelligibility and which specific phonological features they believed contribute to intelligibility. Other questions probed whether the raters' preferred the one-on-one interview or the group-administered format, their actual use of the format,

and their perceptions of the advantages and disadvantages of these two formats. Example

questions were as follows :

What's the purpose of Phase I of the EPT oral section?

What made you decide that the test-taker wouldn't be exempted from the oral
section Phase II?

Which factors of the test-takers' oral performances made you feel that they were
not intelligible?

When you assess the test-takers' oral performances, do you focus on the
recognition of word forms and utterances or the construction of meaning?

More specifically, what issues prevented you from understanding what test-takers
were saying or made test-takers' oral performances unintelligible?

For example, some small pronunciation issue causes confusion
Some words mispronounced
Pronunciation difficulty
Pronunciation occasionally unclear
Sometimes pronunciation is not clear, in particular at word onsets
others

Could you give me some specific examples of problematic words, phrases, or
sentences the test-takers uttered?
If segmental factors caused problems in your understanding, what specific
phonemes made understanding difficult? For example, vowel length, or individual
consonants?

If suprasegmental factors caused problems in your understanding, what specific
aspects such as lexical stress, intonation, or the relative duration of weak and
strong syllables made understanding difficult?

If your decision about test-takers' oral performance in the Phase I was different
from the other TA's assessments in the Phase II, what would cause this
difference?

Do you think the oral group discussion format is fitted to measure the construct
that the EPT oral section assesses?

What do you think are the advantages and disadvantages of the oral group discussion format in terms of the construct the EPT oral section intends to measure?

What was your specific role in Phase I, for example, observing the test-takers' discussion, or conducting the one-on-one interview?

Which format of oral tests do you prefer, and why?

IV. Results

4.1 Intelligibility

The Phase I of the English Placement Test (EPT) oral section adopts intelligibility as its central measure. On the other hand, the Phase II is more concerned about the accuracy. In the first phase of the oral section, the interviewer assesses three minutes of a candidate's speech and seeks evidence of any unintelligible word said that is spoken. If no misunderstanding is detected, the test-taker is exempted from taking the ESL pronunciation course. On the other hand, if a rater doesn't understand some word the test-taker uttered, then the test-taker is sent on to Phase II. Raters in the first Phase also have the option of obligating the test-taker to enroll in the pronunciation course, and they are told that they can do so in the Phase I training.

4.1.1. The construct of intelligibility

Even though the above description is stated in principle, this has never occurred in practice. In general, all the raters seemed to have a common understanding of the purpose of Phase I of EPT oral section. They agreed that Phase I of the EPT oral section is for deciding if the test-takers need to take the ESL pronunciation course or not (even though direct placement into that class does not occur in practice), or in case the raters are not sure if a test-taker needs the ESL pronunciation course, he or she is asked to take

the more fine grained pronunciation test in Phase II. Two of the raters described their purpose and process as follows:

The main test is to put people into 510 or not, and sometimes that's a difficult decision. But sometimes with people it's really easy. It's like clearly you do not belong in 510. So it's kind of to move those people through quickly so that you don't have to take time. So to identify people who are way above or way below the middle. (Rater 2)

The purpose of it, the purpose of the interview is to sort of determine if there are people that you can eliminate right off the bat. After talking for about 3 minutes, their speech is completely intelligible. There might be, they might stumble with words or grammatical things, but you're just looking at the pronunciation, can I understand every word I'm saying, or are there things that are unclear? (Rater 7)

There were no obvious differences of understanding the purpose of Phase I among raters with diverse backgrounds. However, they seemed to understand the construct of intelligibility in rather different ways. Raters' differential backgrounds did not predict these different understandings. Only one of the experienced raters defined intelligibility strictly as word and utterance recognition, which was defined in the same way by Derwing and Munro (1995, 1997, 2006) and corresponds to the definition of intelligibility used by Smith (1980).

I wholly focus on pronunciation. And within that, whether or not there are ever times when I don't understand a word. So I don't care at all about their grammar, their fluency. …..it's also, do I ever have to stop and go, what was that word? So if I have no problem in communication in terms of understanding the word, they move on. But if there's even just a couple of words that I had to take extra time to think about, then I move them through to the second part……I would

differentiate between understanding like a general message, and understanding every word. So it's possible that I'm having difficulty comprehending the main point of what they're saying, but I can understand all of the individual words. That person would, for me, would pass the EPT oral. Because it's, for me, the purpose of that test is, do they belong in 510 or not. And 510 is, does not work with grammar, anything. (Rater 2)

Two experienced raters also adopted the definition of word and utterance recognition as a principle, but they tended to partially employ the notion of comprehensibility, which was defined as judgment on a rating scale of how difficult or easy an utterance is to understand by Derwing and Munro (1995, 1997, 2006). Some researchers used intelligibility and comprehensibility interchangeably (Gass & Varonis, 1984). However, these two constructs are independent to some extent, even though they engage inextricably (Derwing & Munro, 1997). In that study, when examining the relationship among intelligibility, comprehensibility, and accented-ness, Derwing & Munro (1997) found that comprehensibility ratings were harsher than actual intelligibility scores. Rater 3's response echoes this finding:

Well, I kind of go on the guideline provided by the test supervisor that, you know, if, are you understanding everything the person is saying. And if there are words that I'm not catching, if my listener effort is high, then I think, well, maybe that person should go on to the second stage. So my main criteria is am I understanding everything that the person is telling me not…….As far as I know, the guidelines are specific to pronunciation. So their grammar could be terrible, but if I'm understanding the words that are coming out, it might make a difference in how I rate them. You know, intelligibility, yeah. I mean, grammar is tied into that. So if they're really having trouble at the grammatical level, I don't know how much of that is related to pronunciation, or I don't know how much pronunciation instruction would help them with these grammatical issues. And, yeah. I think it affects intelligibility, definitely, if someone is speaking ungrammatically. Listener effort is

high. You may understand all of the words that are coming out, but in terms of how they relate to each other, you have to ask questions to try to find out the relationship between ideas or words or subjects and verbs. So definitely it's hard. Maybe there should be an aspect of the EPT related to specifically to grammaticality. (Rater 3)

Two experienced raters (Rater 2 and Rater 3) assessed the test-takers' utterances as acceptable according to the guideline in the training session if the words and utterances they produced were recognizable and understandable.

Another experienced rater considered test-takers' speech as intelligible when he understood what they were saying in terms of pronunciation, grammar and vocabulary as well as appropriate sociolinguistic use. Thus, when he grasped the overall meaning of the test-takers' speech, he thought that test-taker was intelligible. He focused both on the recognition of words and utterances and the construction of meaning. Following is an excerpt from this rater's interview:

INT: What made you decide that the test-taker wouldn't be exempted from the oral section Phase II?
Rater 4: They receive a 5 (and pass quickly) if they have what I consider near-native understanding and speaking abilities. That means I understand every word, grammatical mistakes are very few and not severe, an accent can be present but is not too distracting.

They receive a 1 (and fail right away) if they have a very hard time understanding my simple questions, if they have a very hard time expressing themselves, if they make extremely bad grammar mistakes consistently, or if their accent is so distracting I miss many of their words.
INT: Which factors of the test-takers' oral performances made you feel that they were not intelligible?

Rater 4: Language that is so garbled or mangled due to incorrect production of vowels and consonants. Incorrect word stress. No rhythm or linking in their speech. Use of unusual, rare, or inappropriate vocabulary.

INT: When you assess the test-takers' oral performances, do you focus on the recognition of word forms and utterances or the construction of meaning?

Rater 4: I focus on both. In phase One, the task is simply to place them into one of three categories, Fail, Continue, and Pass, so there is enough time to step back and make an overall evaluation of the subject's strength as a speaker. For this I include the recognition of words and how they construct meaning and interact with me. (Rater 4)

Less experienced raters understood the construct of intelligibility as word and utterance recognition at least under the situation of Phase I of the EPT oral section. They focused on pronunciation, since they were asked to follow guidelines which ask if they understand every word the test-takers produce. However, in some cases they admitted that inappropriate grammar affected them, or the occasion when they couldn't understand the candidate's topic. Some of them commented that in some cases they define intelligibility in the real world differently:

Um, well. The test supervisor told us to, if we could understand every word they said and not have to stop and think, what was that word again. Then that means that we could understand them well....I'm strictly listening for pronunciation, really, at that point. So even if they're talking about their major and I have no idea what they're talking about because it's crop science or something, as long as I can understand the words they're saying and they're no throwing in a word like crop science, cat, and it kind of makes sense. Then I would pass them. (Rater 8)

I think intelligibility would sort of be, it's like, it's sort of a rating scale. So I would say it's a continuum. You know, not yes or no, but how intelligible is this person? How difficult or easy is it to understand what they're saying? So it's sort of like a scale.... The grammar doesn't affect it. But I do think in the real world, as far as intelligibility, that people who haven't studied ESL and

aren't sympathetic to that maybe, do judge grammar maybe more harshly. Grammar and accent, I would say are two things that like, this person clearly hasn't been in this country a long time. Or they start to stereotype. But with Phase I, you're only looking for can you understand the words they're saying. (Rater 7)

For me it was, no, no. I mean, just like what I would consider that is how much effort I had to put into understanding them. If I could just listen like I listen to most anyone speaking, I guess most native speakers speak English, without really trying to understand what they were saying, then to me that was very intelligible, I guess. And then, I guess, on the other side, if I really had to try hard either to understand what they were saying due to like individual sounds that were unclear, or their overall organization and stress and rhythm, then to me that was less intelligible or unintelligible. I guess they're intelligible in terms of their words and speaking English and the pronunciation is intelligible that you can understand that. But it's probably not coherent in the sense that it's not organized, and whatever they're saying maybe is not, they're using like direct translation, so the phrasing isn't quite right, or. Yeah. I guess that would probably be it. So in terms of intelligibility, it would be easy to understand the words but it isn't really a coherent message. It's not supposed to. According to how it is supposed to go, you're just listening to their pronunciation. We even tell them, like "I'm not listening to your grammar, so don't be nervous about that." But to be honest, it is hard to let, to let it not influence you. Because when you think of someone as proficient or fluent, you kind of wrap all of those things in one package. But technically you're just supposed to listen to pronunciation. So I did my best to just listen to, okay, can I understand what they're saying? It isn't necessarily that they're using good grammar or not, but the clarity of each word or each message is more or less clear. So I tried to focus on that. (Rater 9)

A foreign accent is a common, normal aspect of second language learners' speech, but the impact of an accent on intelligibility is very complex and diverse. An accent has been reported to have adverse consequences for the second language speaker because listeners might have difficulty in understanding speech that differs from local, familiar

patterns of oral production. While some research findings show that accent and intelligibility are closely related, Munro and Derwing (1995) argue that a strong accent does not necessarily preclude fully intelligible speech. In this study, most of the raters believed that strong accent didn't necessarily result in a loss of intelligibility even though it did have some effect on the intelligibility score. One experienced rater (rater 2) said that a test-taker's heavily accented speech did not necessarily affect the score while another experienced rater (rater 4) maintained that a strong accent influenced his understanding of what the test-taker was saying, so it should be considered in assigning the score. Regardless of rating experience or their own language backgrounds, other raters reported that they try to compensate for their familiarity with specific accents.

When I passed the student it doesn't mean that every word that they uttered was perfect. It just means that, I mean, they have an accent, probably, and it just means that I didn't have trouble understanding the word. It was very easy and effortless for me to understand every word……I mean, because the people who gets 5s, it's not as if they all sound American. It just means that their pronunciation in no way hinders communication. (Rater 2)

Research findings about familiarity with specific accents have been mixed. Raters in this study-- in particular the experienced ones -- said that a strong accent was not part of the construct of Phase I, thus they tried not to be affected by the test-takers' accents. However, raters (other than rater 2) admitted that a strong accent was liable to have an effect on scores even though they did not realize it. Such familiarity was related to exposure to a specific accent through teaching and rating experience rather than their own language backgrounds.

Rater 2, an experienced native rater, said that a strong accent did not affect the intelligibility or the scores she assigned to test-takers. On the other hand, another experienced native rater thought that a strong accent could result in a loss of intelligibility, and thus would be reflected in the score. However, he also thought that familiarity with a specific accent could prevent appropriate measurement of intelligibility.

They receive a 5 (and pass quickly) if they have what I consider near-native understanding and speaking abilities. That means I understand every word, grammatical mistakes are very few and not severe, an accent can be present but is not too distracting…if it's uh…foreign language… if it's uh…foreign language, speak a English learner, and they speak not very well, but the listener is very very used to their accent, you know somebody is speaking to Korean people for many many years, then they may not be a accurate judgment of intelligibility.
If some has very heavy accent, then it could affect the intelligibility……Maybe I give higher score to people with familiar accent, but since I know about myself, I tend to compensate, but uhm… like this summer, was the first time I ever spoke to Kazakhstan people, and I spoke with before a first, it was very strange, because I didn't know their pattern, but then after two months, I could understand how they speak, and so I was, you know, more used to it. Since I am an English teacher, that came quickly, but uhm…Yah Yah I can't tell I never test whether or not I rate some reliably. (Rater 4)

Other raters, irrespective of their experience or their native language, were intuitively concerned about familiarity with the accent. Since most of them believed that a strong accent would be a construct-irrelevant variable when assessing intelligibility, they tried not to be affected by their own familiarity with particular accents.

People are more, like for example, like I said, I'm more familiar with Korean maybe, or Japanese accent. Maybe less familiar with Indian accent or Eastern European. So my unfamiliarity might cause me to have those people pass through to the second rating. I'd say that's possible. (Rater 3)

Raters who graded the students in the oral section of EPT for the first time admitted that sometimes accents interfered with understanding what non-native speakers were saying, in particular when the accents were not familiar to the raters, and when the words were spoken with wrong accents. Rater 9 explained it this way:

If it interferes, I guess, with, with understanding with they say. Like certain, I guess, qualities of accents or certain accents, I think, are easier for English speaker to understand, either because they're more familiar with them or because it doesn't affect meaning as much. But if an accent, for example, causes a word to sound different, like for example, errors that often happen with "l" and "r", if it creates a different word, and so there's an error in meaning, then it does. Then it is more of a problem, and more an error in intelligibility. Whereas if a Spanish speaker says, "seet" instead of "sit" within the context, you can kind of understand what's going on, even though it sounds like a different word. So I guess the accent only kind of factors in for me when it changes the meaning of what they're saying. And also too, if it's maybe an accent that I'm not as accustomed to. Like I've gotten used to Korean students or Chinese students in my classes. But up until this semester, I had not had an Indian student. And so the first time when you're kind of hearing accents, and you're just getting used to it…It is harder, I think, to kind of understand. Because you're not used to the way that they pronounce words……Does my familiarity with it affect their score?...uhm…. I would try not to let it. I mean, just because I could understand what may be the Chinese student was saying, I know it's still correct. So I guess the familiarity factor just makes me understand or not, but it doesn't change my opinion of whether or not it's correct pronunciation. (Rater 9)

Grammar and accent, I would say, are the two things that like, this person clearly hasn't been in this country a long time. Or they start to stereotype. But with Phase I, no, you're only looking for can you understand the words they're saying. (Rater 7)

Rater 8, a native English speaker, has had extensive experience with Chinese and Japanese speakers. This is how she reported the challenge:

I live with a Chinese speaker, because my husband's from Taiwan. And a lot of my students are from China, too. So I'm familiar with that. And Japanese speakers, because I've studied Japanese. I usually have a better time understanding them than say, people from Spanish speaking countries. Sometimes I have difficulty with that. And usually I think, like Turkey and the Middle East, I tended to have more problems understanding them, too….I think my familiarity affected the score. Because I had, I don't know all the time if I really understood all the words clearly, but in my head, because I'm familiar with the accent, it seemed like I did. But if I looked at tapes later, I'm not sure I could actually say, "yes, that was each specific word." I hope so, but…(Rater 8)

4.1.2 Contributory Factors

Raters were asked about the extent to which each area of pronunciation contributes to intelligibility: whether segmentals or suprasegmentals are more essential to intelligibility, and which specific factors within segmentals or suprasegmentals were more important for measuring intelligibility. Experienced raters thought that the area of problems varied according to individuals, and these different areas of problems tended to depend on their native language.

INT: So more specifically, what issues prevent you from understanding what test-takers   are saying or made the test-takers' oral performances unintelligible?

Rater 2: Well, I mean, that's, for me that's totally related to the individuals. I mean, sometimes people don't have any kind of rhythm or intonation that's native like, and so I'll have trouble understanding a word. Maybe they put stress in the wrong place. Maybe they don't put word stress. It's possible that they have sound level issues. Like they have trouble with, you know, "l" and "r," or "p" and "f," and they make mistakes that make the word unrecognizable to me. So it's, I think a lot of different factors that are really just related to the individual.

While the raters perceived that the problems contributing to the loss in intelligibility varied according to the test-takers' language background, the raters believed that in most cases problems were related to segmentals.

So it's definitely, there can be like several sounds, there can be vowels might be pronounced incorrectly. So that, you know, there's common errors that people from different language backgrounds make. So that some vowels are, don't sound like the intended vowel. So you can misunderstand the whole word, because there might be, it might be a minimal pair that, if they use the correct sound, you hear that word, but if they use the incorrect vowel, there's an identical, a nearly identical word that you think, oh it's that word. Um, so, yeah. So just distorted vowel sounds, inaccurate consonant sounds. Inaccurate word stress. And not reducing vowels properly, so that every syllable's pretty equal in length. That can make it hard to process and listen. I would say often consonants are the bigger problem. Because, I don't really know how to give you a kind of a principled reason for that. It's just more my intuition. Because, for example, if somebody switches the "v" and "w," which is another common one. Or "p" and "f" for Korean speakers . . .

Because sometimes what's being switched is so, seems unusual to an English speaker. So a "p" and "f" switch just seems strange. "L" and "r," even native speakers of English will occasionally switch them. I think they have some closer relationship in our brains. But "p" and "f" for a native speaker of English just doesn't, it's just not the kind of mistake a native speaker could make. But even the "t" and theta and "d" and "eth," you know, saying "dis" instead of "this." There are certain dialects in American English where people do that. So I don't think we have as much trouble understanding those kinds of things. But probably some of the palatals, like the "esh" and the "guh" [soft g] and "juh," when those are not done properly, they might cause some confusion. I know the "z" sound for Korean speakers, what's the word. There's an organizational, the "z" sometimes isn't produced right. I know I've worked with students where that, not being able to produce that "z" like design . . . (Rater 1)

A lot of it has to do with maybe segmentals. So consonants and vowels that they're really not producing accurately, and so that's causing me to guess at what word they're trying to use. Intonation, maybe not so much. To me, in my opinion, it's more of segmentals.

It depends on the language group. I remember having some Spanish speakers, and they're, some of their segmental difficulties were problematic with the, for example, the "ja" "ya" distinction. I don't know if it annoys me. I wouldn't say it annoys me. But it's just something that's kind of problematic, because it's frequent. With maybe Korean speakers, with the final "g" at the end, like language "g", change "g." So those are kind of clear things. With Japanese speakers, it might be instead of saying for, saying for, using an "h" sound instead of an "f" sound. So I don't know. To me it's language specific…… I mean, when I think of vowels, I often think of Chinese speakers having difficulty, like for example with the vowel "a" or "eh," distinguishing between those. Maybe "e" "i" distinction; "ou," having difficulty with "ou." So when I think about vowels, I know Chinese speakers, from my experience of learning, that they have a lot of difficulties (Rater 3).

When experienced raters didn't understand what the test-takers were saying on the suprasegmental level, they perceived that their difficulties were related to lexical stress rather than intonation or rhythm. For example, when the test-takers put the stress on the wrong syllable of a word that the raters were not used to hearing, they were not understood. Based on their experience, raters thought that intonation had little effect on intelligibility. And since they were supposed to follow the rating criteria to determine if they understood every word the candidates said, intonation or rhythm would not be considered as included in the construct of Phase I.

Intonation, maybe not so much. To me, in my opinion, it's more of segmentals. And also word stress. If they're really inaccurate with their word stress that causes me to kind of guess and listener effort. At the level of rhythm, again, I'm going out on the guideline of am I understanding every word that they're saying, whether or not their rhythm is good, uh, that's kind of another issue. But am I understanding the words that are coming out? (Rater 3)

Lexical stress. You know, if someone says "obliGAtory" instead of "oBLIgatory," I mean I know that example, but if they say it on a word that I'm not used to hearing, sometimes I just have no idea. But I think also just general phrase rhythm. I mean, if the person's not reducing function words and stressing content words, oftentimes that can just get difficult to listen to (Rater 2)

I think lexical stress is the most important (Rater 1)

However, rater 4 had a different opinion. He defines intelligibility as how well the speaker's intended message is understood. He focused on both the segmentals and suprasegmentals.

Mispronounced sounds and a lack of Native English rhythm make subjects unintelligible. Vowels often cause a lot of trouble. When subjects make the relaxed vowels tense, it takes a lot of work to reconstruct what the appropriate vowel should have been. /v/ instead of /w/, and /b/, /p/ instead of /f/ causes problems, the palatals are tough, the worst is when people have troubles with the final consonants, as is the case with Thai or Vietnamese students. Students who don't put primary phrasal stress in the correct place are hard to understand. They don't indicate contrasts, or ask or answer questions how I would expect (Rater 4).

Less experienced raters replied differently when asked which factors contribute most to intelligibility. Two less experienced raters thought that segmentals and lexical stress are important for measuring intelligibility. One less experienced rater considered lexical stress as most essential and the other less experienced rater put the emphasis on suprasegmentals including word stress and rhythm. They also thought the difficulty the test-takers had varied according to test-takers' language backgrounds, and like the experienced raters, one less experienced rater believed that the suprasegmental factors

excluding lexical stress should not be rated in Phase I since they did not directly affect the pronunciation of the words produced.

I mean, intonation is something that you do test on Phase II. It is on the exam. But I only looked at it as, is the intonation so confusing that I can't tell what they were trying to say. So like, which rarely happened. It was really more, I think, certain consonants and things that, or vowels, that sort of distorted the word. Or stress, word stress, that to me distorted the word beyond recognition. So, and then I would consider it unintelligible. Yeah. I can't really think of . . . (Rater 7)

I think that the lexical stress is more important (Rater 6)

I think a lot of it, the vowels had a lot to do with it. Intonation, for the people that I interviewed, I didn't have intonation as much of a problem. I'm not sure about other backgrounds, if the intonation would be different. But the people that I interviewed seemed to use stress and rhythm and that pretty well, in my estimation. Um, the vowels, some vowels and some like blends, consonant blends at the ends of words would be hard sometimes. Like the "th" sound would come out "s" or something. Or the vowel contrasts, like "e" versus "uh" contrast. Those would affect intelligibility for me (Rater 8)

Probably suprasegmentals, when you're considering the whole message unit, not just the individual segmentals, the individual sounds. Because I think that students have a lot more errors with like linking and with rhythm and stress. And since in English we listen so much for where primary stress is, and we reduce vowels so much, when we don't hear that, I think that kind of messes you up more. Or it's harder to understand more than if like you produce a vowel a little bit off, or if you produce a "b" instead of a "v." Like the phenomena that happen on the suprasegmental a little, I think affect more the intelligibility than on the smaller level. (Rater 9)

Raters had different perceptions about which factors contribute more to intelligibility even though their perceptions were not radically different. Most experienced raters perceived that segmentals and lexical stress were more important than

intonation and rhythm when they measure intelligibility. This result is contradictory to previous findings that advocated the primacy of suprasegmentals over segmentals in contributing to intelligibility. This contradiction might be due to the fact that in Phase I of the EPT oral section raters interpreted rating criteria as determining if they understood every word the test-takers produced. The construct of Phase I is closely related to their interpretations of contributing factors for intelligibility. For example, rater 4 responded differently because he defined the construct of intelligibility to include grammatical correctness and socially appropriateness. Therefore, this rater said that intonation or sentence rhythm was important for intelligibility. On the other hand, the less experienced raters had slightly different opinions. They said that word stress and segmental factors were more important than intonation or sentence rhythm for intelligibility since they care about intelligibility or understandability on a word level.

4.2. Group Oral Format

Phase I of the oral section of the EPT currently employs the group-administered format in part. "In part" here means that the test format is not standardized, and raters adopt a one-on-one interview format or a group-administered format at their discretion. When a rater decides to take the one-on-one format, the rater meets a test-taker individually, but recently the number of students who taking the EPT has increased since the university is admitting more and more international students. Which format each rater uses seems to depend on their own preference and judgment of which is more appropriate

for the number of test-takers and spaces available. Most interview questions concerned their actual use of and preference for the one-on-one interview format or the group-administered format. Questions also probed the raters' perception of the validity of each format to allow them to adequately assess the test-takers' intelligibility.

Generally, experienced raters have used both formats according to the number of students and testing places available. If they had to assess a comparatively large number of students, or if they rated the students in a large room, they chose the group format. They also rated students in a group when they were informed to do it this way. But even though they used a group administered format, it was not the same group discussion format discussed in previous research. In most cases, students were not allowed to interact with other candidates, and they talked to raters in front of the group. That is, the rater controlled the conversation. Some raters basically believed that they used an interview format in a group, or they used both versions at the same time.

I've used both. It just depends on the logistics on how many people have and, you know, typically I use, if I'm given a large group of, you know, 20, 30 people. I've seen somebody take them individually. I tend to just take them as a group and have them individually introduce themselves or ask them questions, follow up questions. So both formats. (Rater 3)

Actually, we're employing the group format. I, sometimes I do them in a circle. Sometimes I call people in one by one. They might be waiting. Sometimes we put in different situations. Sometimes we're in a big lecture hall, and I sit over here. But the people I'm going to test are over there, and I call them one by one. Sometimes I'm in a classroom, and I have people wait outside and come in. Other times, I've had everybody come into the same room, and sometimes they sit in rows, if the class is set up that way. Sometimes we sit in a circle. But I just have them go one by one and tell them to introduce themselves. And based on what they say, I may ask

another question. But it's very much, I just want them to talk. I don't try to participate, like I'm making conversation. (Rater 1)

I definitely do it one by one. For me, I think the only reason that doing the Phase I as a group saves time is I don't have to introduce myself to everyone who walks in a room, right? But the way that I do Phase I is literally, everybody is sitting in a little row. I tell them, say your name, say where you're from, talk to me for 90 seconds about why you're here, and something interesting about you. What you did this morning, I mean, whatever, just to get them to talk. But it's completely individual. They're talking to me, and I just go, boom, boom, boom, down the line. (Rater 2)

Most of the less experienced raters adopted the interview format. Three of them said that they were not informed about how to rate students in a group. They also didn't have much knowledge of what the group oral is.

So what I would do is, they're all out in the hallway, and I explain the directions. I say that we're going to do with approximately a 3 minute interview. I am focusing on your pronunciation, the way that you speak. I don't care about your grammar. I don't care about anything, those other things right now, just pronunciation. I'm going to give you a topic, and I want you to try and talk for the full 3 minutes. So I'm going to try and just listen. If I need to ask another question I will, but hopefully I will just be listening. So that they know that they should be continuing the conversation on their own. It's not like an interview. It's just like I ask you a question and you sort of present. (Rater 7)

Phase I was, well, let's see. We would have a group of names, and we'd take them all to like a classroom, a separate area. And then we'd talk to them, usually one by one in there. I guess, I'm not sure how the other ones did it. But I would sit, usually, away from the other students and call people over to me to talk. (Rater 8)

. So the way I did it was, I had the students sit toward the front of the room, and I called them back to me. I didn't want to sit at the front of the room, because I didn't want everyone watching. So I

sat at the back and I called students back one by one, and I just talked to them for maybe 3 or 4 minutes, until I got enough of an idea of what their pronunciation was like. (Rater 9)

Besides the raters' actual use of format in Phase I, they were asked about which format they preferred. Four experienced raters disapproved of the group format. They were concerned that test-takers might be nervous was and the possibility of imitating what other test-takers were saying. They believed that in a one-on-one interview the test-takers were less nervous and had more confidence.

In terms of preference, I would prefer individual. I like meeting people one-on-one and probing them kind of individually. And I think there's also a social factor involved when you're standing up in front of people. I think that kind of affects how you talk and maybe the accuracy of your production. (Rater 3)

Some people might be more shy or not comfortable talking about themselves in front of people. So that's why I don't really like it as much, because I think it could be very, you could get a misreading of somebody. They might perform more poorly than they really can perform because of nervousness or, you know. (Rater 1)

Since they saw what questions are asking, then the other people will have chance to think about it. Also people are very good at reciting. Except that, for example, they are like… I am from this university, I like the baseball, I am really excited to be here, and they can say that very well, and beyond that the language skill is very low. (Rater 4)

The raters thought that test-takers' nervousness or imitating what the other test-takers were saying would make a difference in their performances. Thus test-takers' performances would not reflect their actual ability. The possibility of imitating what other

test-takers were saying varied individually. Raters also said that test-takers who were asked to talk relatively later might perform better. Those test-takers also seemed more relaxed because they had had time to prepare, which might be harmful to the fairness of the test from both the raters' and the test takers' perspectives.

Um, yeah. I mean, overall I would say the test-takers feel more nervous in the group situation. I'm guessing. I'm trying to think of instances where someone was noticeably nervous in the one-to-one. Maybe not as much. Again, with the one-on-one, I'm thinking. With the one-on-one, it's typically in the afternoon when they've gone through all the testing already, and they're kind of tired, and it's just. But a lot of the group interviews are in the morning, first thing. It may be kind of, just, they haven't warmed up yet maybe. And so that might make a difference, like time of day might make a difference in terms of nervousness (Rater 3)

 I think it, people who go first might be more nervous. They don't have preparation time. The people who go later, middle to later, they've had time to think about what they want to say. And I agree with you. They hear what everybody else says, so that gives them some ideas. Well, they can say the same things. I don't, I'd say I've only done that, the group testing, maybe 10% of the time. I really don't like that. I think the only time I've done it is when, you know, we were told to do it that way.
Some people might be more shy or not comfortable talking about themselves in front of other people. So that's why I don't really like it as much, because I think it could be very, you could get a misreading of somebody. They might perform more poorly than they really can perform because of nervousness or, you know (Rater 1)

In addition, the experienced raters reported that they tended to restrict the topic of the tests in a group situation to self-introduction. This topic is familiar to most of the test-takers, which could enable them to show better abilities than they actually had. It is also easy to imitate other candidates' responses if the same or similar questions are asked to all the candidates. That is, raters were concerned about the possibility that test-takers could rehearse their responses, which would prevent them from assessing test-takers'

actual intelligibility. And again, if chances for rehearsal were not identical for each test-taker, it would not be fair.

It seems that some test-takers, when it's the group format, they've introduced themselves before for other, for their own departmental type orientations. So they kind of have a set format of how they introduce themselves. Whereas on the one-on-one, I don't really, I have specific questions there where I don't let them kind of talk, introduce themselves on their own. I have a question that I ask that may not be necessarily a self-introduction question. Does that make sense? So in this kind of group setting, maybe they can fall back on this kind of self-introduction that they've memorized to some degree, or if they've given it before, possibly. But then I always ask follow up questions that, you know,… (Rater 3)

However, one experienced rater had a different opinion about the fear of copying other test-takers' responses. She said that in Phase I the raters are supposed to focus only on the test-takers' pronunciation, and thought that it would be impossible to enhance the ability of pronunciation by copying others' pronunciation within a very limited time.

One of the things that our test supervisor always tells us to do is ask different questions, so that people can't prepare. But I think that, I believe that because I don't care about fluency or grammar or vocabulary, I only care about their pronunciation. And if they have trouble with word stress and rhythm, they're going to have trouble with it whether they have 20 seconds to prepare or not, you know. Or it can't be improved upon in 3 minutes while you listen to other people (Rater 2)

More importantly, experienced raters emphasized that the purpose of the oral section of the EPT is to assess test-takers' intelligibility and pronunciation. In the group format, some of them had difficulty in assessing test-takers' intelligibility. When a student sat in the back, raters couldn't focus on their pronunciation very well.

In terms of intelligibility, possibly, you know, distance from me, and if they're at the back of the class, that might affect intelligibility. If they have a soft voice to begin with, it might affect intelligibility. Whereas if it's one-on-one conference, you know, I'm there with them. I can lean in. So yeah, I think it affects to some degree whether or not I can catch everything that they are saying. (Rater 3)

Experienced raters generally believed that the construct of intelligibility resides in individual, thus it might be better that the test process be done individually. They also thought that Phase I of the EPT oral section needs to be standardized and raters should be trained.

We never, I have never been trained on how to do Phase I. Sometimes, we do Phase I and Phase II together, so we go into the room, and we're doing one person at a time. And we decide. Do we only do Phase I, or do we go ahead and do Phase II right there? And we decide. Do we only do Phase I, or do we go ahead and do Phase II right there? That's sort of the most common way. But when we do that early morning evaluation where it's Phase I only, I never received any training on doing that. But we're told, I mean, we're told to use the same criteria that we do if we're doing Phases I and II together. So I do that. It's if you can, if you have trouble understanding even just a few words, and it's not, you just try to get them to talk as much as possible so you can get a good speech sample. But in terms of doing, how to do it as a group or, you know, just more the practicality of doing it, I never received training. So I always felt like I was sort of winging it, you know, kind of guessing how to do it. But I just did it the same way as when I do the full interview, or Phase I and II together, where you choose, maybe you exempt someone, and they don't finish the test. Or you choose not to exempt them, and they go ahead. (Rater 1)

I think it's pretty, I mean, it, to me, is, it seems like the standardization of the individual meeting has been done pretty well. In terms of the group meeting, maybe less so. I think this is kind of a newer format, and in terms of having papers or materials to give to graders to tell them what to do, it seems like there is a need, in my opinion. You know, I mean, the calibration sessions that

we go through are calibrations showing one-to-one interviews, not group interviews. So sure, I think having a calibration session where you discuss group interview techniques or problems that you may encounter, or again, having maybe a written format, written criteria. I think that would be beneficial. (Rater 3)

The group format is not very fine grained (Rater 4)

On the other hand, less experienced raters used the individual interview format since they preferred to do so or because they weren't informed about how to do the Phase I in the group format, with the exception of one less experienced rater.

I led a group discussion and asked students to introduce themselves. Then I asked them general questions outlined in the interview guideline. Students were in a group, but individually administered. (Rater 5)

Um, some students listened to what the other students said in some classes. I didn't do that. Just because I thought it was kind of distracting and things. So I had them say it alone. Plus, some people were in a lecture. I was in a small classroom. So, um, I would call the one student. And when we finished, I would say, okay, what is the next name. Can you go out in the hallway and get this person? So I just stayed in the classroom and the students would file in. (Rater 7)

In terms of preference, two of less experienced raters had very different opinions about the group format. While a less experienced rater enthusiastically agreed with the idea of a group administered format, another less experienced rater thought the format didn't work well in Phase I.

The group format saves time, lessens test anxiety, and simulates a classroom setting. (rater 5)

My preference depends on the number of students we have. On large days, I definitely prefer doing Phases I and II separately for practical reasons. On smaller days, I like doing Phases I and II at the same time because I think it might seem like a more natural way of interviewing, a format that students might be more familiar with. (Rater 6)

Three other less experienced raters graded students for the first time in the fall semester of 2010. They used the interview format. They were not informed about the group format. When they were asked about the group format, they considered it as a group discussion format in the same way the previous studies have, which is supposed to include discussion in a group. They thought that the group discussion format would have benefits such as saving time, but they might have difficulties in administrating a group discussion format unless they were trained, and they were concerned that students would be nervous during the group discussion.

. I was afraid that that might make the student who's being, who's talking nervous. So that can maybe affect your pronunciation if you're, not your pronunciation, but just like the fluidity of your speech, if you're very nervous. And also, I thought like, because the prompt should be not something you're expecting. You know, if I wanted to give them all the prompts in the hallway, I would have. But I wanted them to talk naturally. So if you've just heard someone, there were only like 10 questions on the worksheet. And I only ask usually four or five of them. And so if you already heard all those students give a response, you can start preparing your own responses. Then when you're preparing it, then I feel like you have more time to think about, oh, okay. How would I say this? Oh, I don't know how to say that word. How can I get around it? [laughs] And so I didn't want them to have that time to think about, okay, like circumlocution and other strategies to get around it. I wanted it to be sort of a natural, on the spot response. So that's why I kept them in the hallway. (Rater 7)

I think, it does have some advantages, I think. I think the students might feel a little bit more comfortable knowing they're talking with other students too, not just me, who they know I'm

evaluating them. And also for me as far as comparison, if I hear two or three speakers responding kind of at once, then I can kind of see, "oh, he speaks a lot more clearly than he does." Or, "wow, she really needs to go to Phase II, because I'm not understanding her like I am him." But for disadvantages, if the group is talking and they start talking and talking over each other, I have a lot less control of what I can hear and what I can pick up. And even time limits. And if I ask a question, and they say one statement each, and then they just sit there and think like, okay, we're done now. It might take longer. I don't know. But I think that there are advantages to that too. (Rater 8)

Interveiwer: You did any group format? Did you do any group interviewing?

Rater 9:   This, so it was the first time that I had done it. I didn't know if I could, how I would organize the group discussion and like accurately like rate everyone if multiple people were talking. And since the students didn't know each other, I wasn't sure how productive a group discussion would go.

Interviewer: So what could be disadvantage or advantage, or like, do you like that format? And do you think it's going to work well or not? So what's your opinion about group format?

Rater 9:   I see the benefit to it, that it would save time and that it, because there are so many that come through. It would be more efficient in terms of time. I'm just not sure, my doubts with it is that, like I said, I don't, I'm not sure the students would generate a very productive discussion with people they don't know at such an early hour in the day. I can't imagine it being, you know, very talkative. And then at the same time, because it's not my class. It's not like a class of students that I know. I have to be interrupting constantly to be like, wait, what was your name, what was your name? Because it's kind of hard for the rater within 5 minutes to memorize the name, the name of 20 people in the room.

Interviewer: Yeah, uh huh.

Rater 9:   So how do you accurately assign a score without, you know, either having them wear name tags or something. So I guess it would take a little more, maybe planning and practice. We might have to have like a practice run with students that would be willing to try it. But, I suppose it's possible. But it would take maybe some more planning than just saying you're going to interview the whole group and switching right away.

Interviewer: Yeah, so can you, can you imagine a situation where the students can interact to each other, not just with you. So is it, so is it good for your rating? You can rate the students when they interact to each other. So is it, does it cause any problem on your rating or?

Rater 9: Having them interact, like the way that they interact with each other versus?

Interviewer: Yeah. You just watch what they are saying and what they are listening to, interacting things.

Rater 9: Um, I don't think it would really affect it either way. I mean, I would still be hearing their voice, which is the important part. But if you're having, like if you set them up in pairs and they were just having a conversation with one person, you would have so much background noise that it would be hard to hear everyone. And then similarly, I guess, if you had like a whole group, you would have to deal with like the problem of turn taking and interruptions or restarting. And so it's more to get through to focus on their pronunciation. (Rater 9)

The problems experienced in the group format were also predicted by less experienced raters even though some of them hadn't experienced the group format. Like the experienced raters, they were also worried about the possibility of varying opportunities for students to rehearse their performances, the risk of imitating other test-takers' speech, and the various extents to which test-takers would be nervous due to the pressure of having to speak in front of others. They thought that if the test-takers had time to prepare what they are going to say by watching what the others are saying, or if they were nervous about talking in front of the other students, it would affect their pronunciation, and they might not show their real abilities. However, at least the less experienced raters appeared to have a relatively more positive attitude toward the group format than the experienced raters, though they were still concerned about the possible difficulties of administration.

I see the benefit to it, that it would save time and that it, because there are so many that come through. It would be more efficient in terms of time…..I suppose it's possible. But it would take maybe some more planning than just saying you're going to interview the whole group and switching right away (rater 9)

I prefer the group-administered oral because it saves a lot of time. The primary purpose of the Phase I is to save time and I feel that the group oral format does what it is supposed to do. In addition, even if students receive hints from previous students, because the purpose of the assessment is pronunciation, I don't think it will affect the results. If the raters are properly trained, it should not be a major issue (rater 5)

V. DISCUSSION AND CONCLUSION

5.1. Intelligibility

The oral section of the English Placement Test measures intelligibility. It adopts a subjective, word-based definition of intelligibility (Isaacs, 2008). However, this notion of intelligibility is defined differently from related research due to the specific context of the EPT oral section. While intelligibility is operationalized in a single-stage event from the language testing literature, the oral section of the EPT has a two phase structure.

In the first phase, the raters assess test-takers' performances based on their being able to immediately understand every word test-takers utter. The purpose of Phase I of the EPT oral section is to determine if test-takers need to take ESL pronunciation courses. When raters are able to immediately understand every word test-takers say, they are exempted from taking the ESL pronunciation course and from taking the full pronunciation test, Phase II of the EPT oral section. If this is not the case, test-takers take Phase II of the EPT. This full pronunciation test contains a brief one-on-one interview at the beginning, providing test takers another chance to get a score of '5.', which means that test takers are to be exempted from taking an ESL pronunciation class. If raters judge test takers as not intelligible at this interview, then test takers go on to sentence/phrase/dialog reading, during which the rater would make a decision about the need to take an ESL pronunciation course.

Therefore, in the first stage, intelligibility is defined as 'full, immediate understanding'. If raters do not understand even for one word fully and immediately the

interview, they are instructed to continue with the word/phrase/dialog part of the test. They are specially instructed not to try to understand a particular word or to work to comprehend general meaning or to give the test taker another chance to be incomprehensible. One word that is not fully and immediately understood makes raters decide to continue with the word/phrase/dialog part of the test, which is given in the Phase II.

However, it is not the case that the EPT oral exam has always two parts. When the number of test takers is relatively small, the test coordinator may decide there is enough time for all the test takers to take Phase II. In this case only the Phase II is administered. This shows clearly that Phase I functions as an oral screening measure, and its main purpose is to save time and the need for raters.

The first interview questions were about the purpose of Phase I. Raters all said that they intended to measure test-takers' intelligibility or pronunciation in Phase I, and they decided if test-takers need to take ESL pronunciation courses or not.

However, even though all the raters agreed that Phase I assesses test-takers' intelligibility, they defined intelligibility in different ways, and they used different criteria when they assessed the test-takers' speech. Neither their rating experience, their teaching experience nor their language background predicted these differences. How examiners defined intelligibility appeared to vary individually. An experienced rater who had

several years experience teaching ESL pronunciation courses and rating the EPT oral section defined intelligibility as 'word and utterance recognition.' Her use of intelligibility seemed to be equivalent to Dewing and Munro's (1995, 1997, 2006) notion of intelligibility and Gass and Varonis's (1984) notion of comprehensibility in the way the constructs are operationalized. However, as discussed earlier for several times, this operational definition as word recognition is not exactly the same as those from the related research. In this study, the word recognition must happen immediately. If raters do not understand even for one word fully and immediately the interview, they are instructed to continue with the word/phrase/dialog part of the test. Her definition was different from other raters in that she differentiated between getting a general message and being able to recognize every word. She admitted that even though she could understand every word, she might have trouble in comprehending the main point of what test-takers intended to say. In this case, she thought that this person would pass the EPT oral section. She had taught ESL pronunciation courses for many years, so she had ample knowledge of what was to be taught there. She claimed that the ESL pronunciation classes only address pronunciation, thus in the EPT oral section, pronunciation is what should be assessed.

Two other experienced raters who also had experienced teaching and examining students for several years said that the criteria they adopted was understanding every word test-takers utter, and they focused only on pronunciation. However, they admitted that if the main points of messages were not appropriately delivered, it might affect how they rate. Ungrammatical speech also might have some consequences. Their definition of intelligibility seemed to include word and utterance recognition as well as ease of understanding.

Another experienced rater reported considering various aspects such as pronunciation, grammar, vocabulary, and appropriate sociolinguistic use. He believed that all these factors would affect intelligibility. He said that if test-takers did not express themselves effectively, their grammar was not consistently appropriate, or their speech was heavily accented, then he thought test-takers were not intelligible. Use of unusual and inappropriate vocabulary was also a factor of unintelligibility. Thus, his definition of intelligibility entailed word and utterance recognition as well as the construction of meaning. In terms of Smith's (1980) interpretation, this rater's perspective includes all aspects of intelligibility, comprehensibility, and interpretability.

The experienced raters did not all use of the definition of "full, immediate understanding" in the same way even though most of them knew that the guideline they were supposed to follow was to determine if they understood every word. They also knew very well that if they are in doubt, further testing would happen in the Phase II. Less experienced raters, on the other hand, said that they focused only on pronunciation. They commented that sometimes they had difficulty in understanding what test-takers intended to say because of unfamiliar words that were field-specific or when they referred to test-takers' own culture and places in their own countries. However, raters revealed that even though each word or each message was not clearly delivered, they concentrated on being able to understand every word and how much effort they had to exert to understand test-takers' speech. Some of them said that this is the principle they were supposed to follow in the EPT oral section, and beyond the context of the EPT oral section, non-native speakers' intelligibility would be defined in another way, possibly including correct use of grammar and accent that is not too distracting. In addition, a rater

with less experience stated that it is hard to say that severe grammatical mistakes and very heavy accent had no influence on rating.

As we see from the results, all the raters knew that they are supposed to answer the question, "can I understand everything the speaker say?" for the Phase I, and if they are in doubt, further pronunciation testing will occur in the Phase II. Hence, the operational definition of intelligibility for the EPT oral section is that the rater have full, immediate understanding in order to determine if the test taker is intelligible, and if the rater is in doubt, then there is Phase II for the further assessment. But, not all the raters defined intelligibility according to this operational definition. . Their teaching experience and language backgrounds did not seem to have much effect on their interpretation of intelligibility. Various factors appeared to influence the raters' ideas in that regard. Exposure to non-native speakers' speech would be one factor, and the extent to which they had knowledge of what is taught in the ESL pronunciation course would also be a critical factor in differentiating their conception of intelligibility. Most of them seem to internalize their rating criteria according to Smith's definition of intelligibility that focuses on word and utterance recognition. This definition is limited to the phonological features of speech. However, raters sometimes included semantic concerns such as the accessibility of literal meaning or concerns about grammatical ability.

Accent appears to be one of the variables that influences listeners when they rate non-native speakers' speech. The relationship between intelligibility and accent has been examined for several decades. Numerous studies have shown that native speaker listeners tend to downgrade nonnative speakers' speech because of foreign accent. Heavily accented speech has often been considered unintelligent. Even though the view that

accent and intelligibility are closely related has been prevalent among researchers, Munro and Derwing (1995) claim that accent and intelligibility are related but partially independent dimensions based on their study findings indicating that a strong accent does not necessarily preclude fully intelligible speech.

The individual raters in this study did not have identical opinions about the extent to which accent contributes to intelligibility. Most of them agreed that heavily accented speech was not necessarily unintelligible. One experienced rater said that test-takers with an accent would be intelligible if she was able to understand every word they uttered. From her perspective, their speech did not need to sound American. Other raters also agreed that test-takers' pronunciation with an accent, in some cases with a heavy one, might not hinder oral communication. However, one native rater with much experience had a distinctive view about an accent. If a test-taker's accent was distracting, he argued that it could affect intelligibility.

In summary, every rater agreed that familiarity to a specific accent could differentiate his or her ratings. Familiarity with specific accents resulted from exposure rather than raters' language backgrounds. This research finding corresponds to Smith and Bisazza's claim (1982). In their investigation of the factors of comprehensibility, the authors sum up:

It seems clear from this study that one's English is more comprehensible to those people who have had active exposure to it. With English being used frequently by nonnative speakers to communicate with other nonnative speakers, this study gives evidence of a need for students

of English to have greater exposure to nonnative varieties of English. (Smith and Bisazza 1982:269)

Familiarity to specific accents could be resulted from various sources. The raters from the interview stated that not only raters' native language, but also raters' prior L2 learning experience, their living experience in non-English speaking countries, and the extent of their interaction with non-native speakers could affect the familiarity specific accents. Therefore the training session should include various sources of bias related to rater. Even though it is obvious that raters for the EPT oral section should be native speakers or non-native speakers with extremely high English ability, we should note that even native speakers are not absolutely free from these various sources of bias, and native language of non-native speakers are not the only source of bias.

The last interview questions about intelligibility focused on which phonetic or phonological properties explained intelligibility more. Most raters assumed that lots of phonemic or phonological problems varied according to the individual test-taker. Additionally, all of them perceived that there might be common errors that people from different language backgrounds make. For example, raters recalled that Korean speakers of English often switched /p/ and /f/, and had difficulty in articulating /dg/ sound while Chinese speakers of English had trouble in distinguishing between /e/, /i/, and /ou/. Generally, as to segmental factors, most raters tended to consider that problems in consonants were more likely to lead to unintelligibility. While there have been contradictory results regarding whether consonants and vowels are more important for intelligibility and communication, Sewell (2010) and Jenkins (2000) explained that while

only a small range of consonant modifications are accepted in terms of intelligibility, vowel substitutions are treated more leniently. This can be explained from the fact that if the vowels are removed, a listener would be able to decipher the word from context, but not if the consonants were removed (Sewell, 2010). However, raters in the EPT oral section made comments only about difficulties from different language backgrounds and didn't specify which phonological features actually caused more difficulties in intelligibility. Most raters recognized that word stress is one of the most important features contributing to intelligibility. This might be due to the fact that Phase I assesses intelligibility on a word level and thus raters are supposed to measure if they understand every word test-takers utter. For these reasons, word stress was considered more important than sentence rhythm or intonation.

Another interesting result is that experienced raters had different rating criteria than less experienced raters in assessing test-takers' problem areas. Even though test-takers made multiple mistakes, there were some occasions where those mistakes were limited to one problem area such as the /p/ and /f/ difference. In this case, a test-taker was likely to pass Phase I. This is due to the fact that raters thought that this test-taker would pass Phase II because there is only one item related to this problem area in the questions asked in Phase II. Since experienced raters had enough knowledge of what would be assessed in Phase II and what would be taught in the ESL pronunciation course, they could decide what they should focus on during Phase I.

5.2 The Group Format.

Phase I of the EPT oral test uses a group-administered format in part. Since a huge number of students are taking the EPT recently and graders have to determine if test-takers would pass, go through the second phase, or fail the first phase within a very limited time, raters are encouraged to employ the group administered format. The group format that the EPT employs is not the same logistical situation as discussed in the research examined. In the research reviewed for this thesis, group oral testing allows interaction between examinees, and in addition, the result of the single exam is a single and final score. In the EPT Phase I, however, the group setting is a practical convenience, and raters know that test-takers who are not exempted will go on to a one-on-one assessment in Phase II. For this reason, it is probably best if a new term is introduced to categorize Phase I of the EPT, so as to distinguish it from the growing literature on group oral exams: a good candidate would be to call it a "grouped" exam, using that adjective to mean, simply, that the test-takers were grouped together out of logistical convenience. Although the EPT Phase I was originally designed as a "grouped" test, it appeared that all the raters did not use the same format. Experienced raters employed both the interview format and the grouped format at their discretion. They decided which format to use depending on the number of test-takers and places available. On the other hand, three of less experienced raters employed only the interview format. They were not informed of which format to use. Raters reported that since they did not all use the same format, and they were not appropriately trained for the format, this could have an effect on their scoring.

Most raters preferred the one-on-one interview format in Phase I. Only one rater with less experience preferred the grouped format to the interview format since it saves time.

Other raters disagreed with the grouped format for several reasons. They believed that several variables rather than test-takers' abilities could affect their scores. First of all, raters thought that the grouped format didn't provide an appropriate context for assessing intelligibility. Under the grouped format, factors other than test-takers' pronunciation could differentiate the scores test-takers received. Distance from raters, or test-takers' voice quality would be reasons for reducing intelligibility. Furthermore, raters were concerned that when they had to grade students in a group, they were easily distracted, and had difficulty in focusing on individual test-taker's pronunciation. Analysis of experienced raters' responses indicated that intelligibility or pronunciation ability would reside in individuals. Thus from raters' viewpoints one-on-one interviewing would be a better way to assess an individual's intelligibility rather than a grouped approach.

Another issue raters raised was ordering effect. They were concerned that there might be a temptation to copy another test-taker's response. When test-takers talked in a group, they were asked several questions by raters one by one. Raters thought that test-takers who were asked to talk relatively later might perform better. They could have time to prepare, assume what they would be asked, and have the opportunity of repeating and reconstructing what other test-takers were saying. This advantage seemed to be against the rating criteria. Thus, it would be harmful to the test's fairness as well as its reliability. However, there were different opinions about the possibility of copying other test-takers' responses. One experienced rater and one less experienced rater believed that it was

unlikely that test-takers would benefit from copying other test-takers' responses. Since Phase I basically assesses candidates' pronunciation rather than the content or overall grammatical structure of their speech, it would be difficult to fix their pronunciation within that very limited time. Even though candidates had the possibility of rephrasing or mimicking parts of other candidates' speech, it would not affect on their scores.

Additionally, raters commented that with a grouped format they would restrict the topics of test-takers' performances. Experienced raters said that they tended to depend on self-introduction questions in a group setting while they did not rely on any specific or pre-determined questions with the one-on-one interview format. They assumed that answers to self-introduction type questions were easily memorized, and it was more likely that test-takers would have been asked the same or similar questions before taking the EPT. Therefore they could have had chances to rehearse these questions, and they might perform better than their actual abilities.

Raters also considered that students might get nervous or feel uncomfortable when they had to talk in front of other students. Experienced raters recalled several situations where they noticed that students were embarrassed in a grouped format, making them misread the abilities of the students. They felt that test-takers who were more introverted would be likely to have more trouble speaking in front of other people. The fact that personality variables on grouped tests were noticed by raters is remarkable. In a series of research on a group oral test, there is some evidence that personality factors such as assertiveness, shyness, or extroversion can affect one's own score as well as the scores of the members of one's group (Berry, 2004; Ockey, 2009). Berry (2004) found that it was advantageous to be grouped with extroverts since they were active speakers

and spent a longer period of time discussing an assigned topic. Ockey's (2009)-study also indicated that the personal characteristics of a test-taker's group members could affect a test-taker's score on the group oral. Ockey (2009) found that assertive test-takers received higher scores when grouped with non-assertive test-takers and lower scores when assessed with other assertive test-takers. This personality factor seems to function in a grouped setting. A grouped format in the current study is distinct from the group oral test in that no interaction between test takers, non-native speakers, is hardly acceptable and the grouped format here intends to screen the test takers for the next actual placement oral test. Nevertheless, since test takers are supposed to perform in a grouped setting, introversion/extroversion might be even more of an issue, and raters in this study concerned that some test-takers might perform more poorly than they do with one-on-one interview format. Therefore, research on how such personality variables in a grouped oral test are related to test takers' performances is suggested.

Experienced raters and two of the less experienced raters have employed the grouped formats in Phase I and are aware of several problems. The rest of the less experienced raters who had no experiences using a grouped format thought that they would have problems similar to those of the experienced raters.

The grouped format was considered inappropriate for the oral section of the EPT by most raters. It had the advantage of saving time, but because the EPT oral section mainly measures an individual's intelligibility, assessing test-takers individually was regarded more proper. The risk of imitating other test-takers' speech and the possibility of underperforming when having to speak in front of other test-takers were also thought to have negative consequences on test scores.

Finally, it is noteworthy that some raters did not seem to understand the original intent of the test. As discussed above, Phase II was supposed to entail a brief one-on-one oral interview at the beginning, where raters could award a score of '5,' a passing grade. However, in particular when Phase I and Phase II occurred on the same day, some raters only had test-takers read aloud phrases and sentences in Phase II. That being said, for these raters, it was likely that they in effect interpreted that only Phase I measured intelligibility, not Phase II. However, as noted before, both Phase I and Phase II were designed originally to assess 'intelligibility' under a less controlled setting such as a conversation or interview. Raters' responses revealed that they did not interpret the original intent of the test in the same way.

5.3. Suggestions

The results of this study show that even though raters recognized very well from the training session that they must have full, immediate understanding in order to determine if test taker is intelligible, and if the rater is in doubt, then the test taker is sent to Phase II, they defined this guideline from the training session in different ways, which resulted in their employing different criteria when they made a judgment on the L2 English speaker's pronunciation and intelligibility.

Therefore, the results suggest that the definition of intelligibility needs to be clarified and unified. According to Nelson (2008), while Smith (1980) differentiated strictly between intelligibility and comprehensibility, Munro et al. (2006) conflated processing sound and meaning in the notion of intelligibility. In this study, raters'

responses also showed that they interpreted intelligibility somewhat differently. In particular, less experienced raters understood intelligibility as both "actual understanding of speech production" and "listener's difficulty in processing what they have heard." The former understanding corresponds to intelligibility, and the latter points to comprehensibility in Munro et al.'s study (2006). It is important to note that the purpose of this test is to place students into appropriate pronunciation classes, which needs to be distinguished from general speaking courses or communication classes. Therefore it might seem reasonable that in this test, test-takers' pronunciation should be prioritized over their interactional competence or grammatical abilities. When considering alignment with the content of ESL pronunciation courses and the original intention of the test, intelligibility might be better defined as understanding every word the test takers utter, which would be close to "actual understanding of speech production".

Nevertheless, it is not safe to say that the only source of deviation in raters' different criteria comes from the recalibration training. It might be the case that the rater decided to use different definition from the guideline given, namely "full, immediate understanding". Their own teaching or rating experiences would cause them to interpret the guidelines on their own subconsciously or consciously. Other factors than the training session might inadvertently influence the ratings the raters assign. Moreover, even if the raters show variation in their grading in Phase I, the EPT oral section has the second phase, where further assessment significantly mitigate or lessen ~~these~~ this variation.

In addition, the purpose of the recalibration session did not appear clear to every rater in one more aspect. Experienced raters realized that even though test-takers made errors in one problem area, it might be better to pass those test-takers since errors in one

problem area would be considered as only one error in Phase II and would not trigger a course requirement. Therefore, raters could be better trained in how to deal with the types of errors that candidates make. If candidates produce unintelligent pronunciation in the same area, it would be better to count these errors as one problem area.

As the findings show, raters' language backgrounds did not have a strong influence on their decisions. However, raters' amount and type of exposure to the candidate's accent did influence, either positively or negatively, their rating. Native raters and non-native raters had various amounts and types of exposure to specific second language learners' accents. These findings suggest that non-native raters should not necessarily be excluded from the pool of raters for the EPT oral section if they are well trained in the recalibration session.

Lastly, but not less importantly, all the raters do not necessarily use the same format. The raters in this study used two different formats at their discretion: interview or grouped. Experienced raters employed the specific format they preferred or considered more appropriate. Most of the less experienced raters used the one-on-one interview format since they were not informed about the grouped format. The grouped format had several consequences on scores such as test-takers' nervousness and the fear of imitating other test-takers' responses. Thus, it might seem fairer for every rater to use the same format.

The grouped format is considered to have both positive and negative effects on the candidates' scores. One of the advantages of using the grouped format is being able to assess test-takers within a limited time. If the grouped format is chosen for Phase I, it might be better for raters to be trained to administer the grouped format more properly.

Training should include learning to create a comfortable atmosphere in which students feel less nervous and to diversify the topics of conversation in order to minimize the chances of imitating other test-takers' responses.

Finally, exposure to specific accents appeared to have consequences on raters' judgments. Therefore, if raters could have opportunities to listen to foreign accents that they are rarely exposed to, for example, Arabic accents in the training sessions, it would be helpful for them to make more balanced judgments when grading second language learners' pronunciation.

5.4. Limitations

This study has some limitations. The study results are entirely based on the raters' replies. It might be better to investigate in depth how raters' different understanding of intelligibility affected the test-takers' scores using the data of test-takers' scores.

Another limitation that must be discussed here is that I analyzed the data ~~about~~ on the group format ~~from~~ using both experienced raters and less experienced raters when only the experienced raters had used the group format. This could compromise the reliability of the results. The simple comparison of replies from both experienced raters and less experienced raters could be problematic since they were not in the same situation.

In addition, even though raters maintained in their replies that their language backgrounds or native language had little influence on their grading, analyzing the possible differences of their grading according to their language backgrounds or different

exposure to foreign accents would yield more concrete answers to this inquiry. To better determine the validity and reliability of the two formats considered, a comparison of scores from the one-on-one interview format and grouped format would enable us to find the differences between them. Similarly, raters self-report phonological factors they think more important for grading. However, there is no evidence that in the actual testing these phonological factors contributed more to their grading.

This study also does not investigate how test-takers understand the purpose of the test, whether they think their language backgrounds affect their scores, and how differently they perform with the individual interview format and grouped format. Further research is suggested to analyze data from test takers. This would offer a better understanding of how the structure of the EPT oral section affects test-takers' performance and scores.

REFERENCES

Abercrombie, D. (1949). Teaching pronunciation. *English Language Teaching, 3,* 113-122.

Anderson-Hsieh, J., Jhonson, R., & Koehler, K. (1992). The relationship between native speaker judgment of nonnative pronunciation and deviance in segmentals, prosody and syllable structure. *Language Learning, 42*, 529-555.

Bachman, L. F. (2001). February: Speaking as a realization of communicative competence. Paper presented at the meeting of the American Association of Applied Linguistics – International Language Testing Association (AAAL-ILTA) Symposium. St. Louis, Missouri.

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

Bamgbose, A. (1998). Torn between the norms: innovation in world Englishes, *World Englishes, 17,* 1-14.

Bansal, R. (1969). *The intelligibility of Indian English*. Hyderabad, India: Central Institute of English.

Berry, V. (2004). A study of the interaction between individual personality diffrences and oral perdormance test facets. Unpublished doctoral dissertation. King's College. University of London.

Bonk, W. J. & Ockey, G. J. (2003). A many-facet Rasch analysis of the second language group oral discussion tasl. *Language Testing, 20,* 89-110.

Bonk, W. J. & Van Woere, A. (2004, March). L2 group oral testing: The influence of shyness/outgoingness, match of interlocutors' proficiency level, and gender on individual scores. Paper presented at the annual meeting of the Language Testing Research Colloquium, Temecula, California.

Chalhoub-Deville, M. (1995). Deriving oral assessment scales across different tests and rater groups. *Language Testing, 12*, 16-33.

Chalhoub-Deville, M. & Wigglesworth, G. (2005). Rater judgment and English language speaking proficiency. *World Englishes, 24*, 383-391.

Derwing, T. M., & Munro, M. J. (1997). Accent, intelligibility and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 19*, 1-16.

Derwing, T. M., Munro, J. M., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 18,* 393-410.

Fayer, J. M. & Krasinski, E. (1987). Native and nonnative judgments of intelligibility and irritation. *Language Learning, 37,* 313-26.

Folland, D. & Robertson, D. (1976). Towards objectivity in group oral testing. *English Language Teaching Journal, 30,* 370-167.

Fulcher, G. (1996). Testing Tasks: Issues in task design and the group oral. *Language Testing, 13,* 23-51.

Hadden, B. L. (1991). Teacher and nonteacher perceptions of second-language communication. *Language Learning, 41*, 1-24.

Hildson, J. (1995). The group oral exam: advantages and limitations. In Alderson, J. and North, B., editors, *Language testing in the 1990s: the communicative legacy.* Hertfordshire: Prentice Hall International, 189-97.

He, L. & Dai, Y. (2006). A corpus-based investigation into the validity of the CET-SET group discussion. *Language Testing, 23,* 370-402.

Hildson, J. (1995). The group oral exam: Advantage and limitations. In Alderson, J. & North, B., editors, *Language testing in the 1990s: The communicative legacy* (pp. 189-197). Hertfordshire: Prentice Hall International.

Isaacs, T. (2008). Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *The Canadian Modern Language Review. 64,* 555-580.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford. England: Oxford University Press.

Kenworthy, J. (1987). *Teaching English pronunciation.* Oxford. England: Oxford University Press.

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

Morley, J. (1994). A multidimensional curriculum design for speech-pronunciation instruction. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New views, new directions* (pp. 66-91). Alexandria, VA: TESOL.

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehension and intelligibility in the speech of second language learners. *Language Learning, 45*, 73-97.

Munro, M. J., Derwing, T. M., & Morton, S. L. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition, 28*, 111-131.

Nevo, D. & Shohamy, E. (1984). Applying the joint committee's evaluation standards for

the assessment of alternative testing methods. Paper presented at the Annual

Meeting at the American Euducational Research Association, New Orleans,

Louisiana. (ED. 243 934).

Ockey, G. J. (2001). Is the oral interview superior to the group oral? *Working Papers on

Language Acquisition and Education, International University of Japan, 11,*

22-41.

Ockey, G. J. (2009). The effects of group members' personalities on a test taker's L2

group oral discussion test scores. *Language Testing, 26,* 161-186.

Schairer, K. (1992). Native speaker reaction to non-native speech. *Modern Language

Journal, 76,* 309-319.

Shohamy, E., Reves, E. & Bejarano, Y. (1986). Introducing a new comprehensive test of

oral proficiency. *ELT Journal, 40,* 212-220.

Smith, L. E., & Nelson, C. (1985). International intelligibility of English: Direction and

resources. *World Englishes, 4*, 333-342.

Smith, L. E., & Ralfiqzad, K. (1979). English for cross-cultural communication: the

question of intelligibility. *TESOL Quarterly, 13*, 371-380.

Van Moere, A. (2006). Validity evidence in a group oral test. *Language Testing, 23,* 411-

440.

Van Moere, A. & Kobayashi, M. (2003, July). Who speaks most in this group? Does that

matter? Paper presented at the annual meeting of the Language Testing and

Research Colloquium, Reading, UK.

Rater 1

INT: So, What is the purpose of Phase I in the EPT oral section?
R1:   Ok, So, my understanding is that a way of uhm… more quickly identifying who needs to take the test more the version of the test.
INT: So, Do you think that the Phase I can be a separate measure… just complimentary measure for the Phase II?
R1:   well… It seems like it's identifying who's at the extreme. So, it's sort of separate measure. That's you identify who's clearly a 5 and exem… probably exempted and who might be one. Somebody who really should be required to take a test, but it really can't distinguish among 2s, 3s, and 4s.
INT: Yeah, so, which factor of test takers' performances do you assess?
R1:   Uh… their pronunciation. Because the test is used to screen which student should take ESL 510 and the purpose of that class is to help the pronunciation…is really…I am not focusing on communicative ability that more…how…well I can understand what they are saying.
INT: So, what mainly decided the test taker wouldn't be exempted from the Phase II?
R1:   Uhm… Usually, the main guideline is if their… even if there's a word or two that you can't understand. But, because I 've taught ESL 510 and other courses for pronunciation, I know what… what they learn at a class. And it's usually…somewhat really there's sound errors…serious problems with uh… rhythm that interfere with their being understood… So, it has to be…uhm… I just imagine them talking to native speaker who has never spoken with someone with an accent, and how much trouble they would have.
INT: Yeah, it would be a similar question but, then which factors of test takers' performance made you feel they are not intelligible?
R1:   they are not intelligible? If if they ..the word is a key word is …I can't understand it, so I don't know what they are talking about.. and usually the reason I can't understand the word is because the sounds are incorrect, a word stress is incorrect, so, I might not be able to understand the name of a town or their major, or, uhm… you know some… uh… they are talking about their hobby, so If I can't understand a key word, uhm… uh… so it might be something like that, or some speakers have very distinctive rhythm, to their speech, so different from English that, uh… no… that is hard to identify where one word begins and one ends … their sounds might be pretty good, but the rhythm is different that I can't understand …
INT: So, more specifically, what factors do you think more contribute to the intelligibility? I mean in terms of segmental factors or suprasegmental factors? From your experience?
R1:   well, I've read probably some of the related articles, uhm…. I think they interact, sometimes one is more serious than the other. For example, someone had serious problem like l and r, those are big, you know, a big problem area. And everytime they are using the wrong sounds, I think that could really cause a large trouble, even if their suprasegmental is good. Uhm , I think it is sort of extremes. If…if… one of them is particulary not native like, if their rhythm is like Indian English, their rhythm is very different from American English, so it's very hard to understand. Even their sounds are perfect. So I think it is sort of the extremes

somebody can have really poor sound level problems, but not answering your question. Uhm… usually, the rhythm is the bigger problem the supersegmentals, I think.

R1: So your question was, what I'm listening for, meaning or sort of the phonetic?

INT: Yeah. So you focus more on the like phonetic words, just for the oral section, for EPT.

R1: Mmm hmm, yes.

INT: Then like, so could you tell me some specific examples of problematic words or phrases or sentences the test takers have uttered from the previous EPT oral sections? Can you remember some problematic words or phrases?

R1: Oh, let me think. Probably not specific. I think the most common, maybe I can give you more categories of problems rather than. I don't, I don't think I remember specific words or phrases. But for example, let's say it's somebody, a common problem is that people learn to read English first. So when they come here, they're trying to speak, you know, without having practiced much speaking out loud. So they're kind of going by their interpretation of what they, they know how to read, which they do very well. So often it's just people will leave off syllables at the ends of words, or they'll use incorrect word stress, so it changes the rhythm of the word, which makes it hard to identify it. And then other times they'll be just serious sound errors. The "l" and "r" I mentioned. A Vietnamese speaker will say an "s" instead of an "esh" or will leave off actually all the sounds at the end of syllables. So that makes it really difficult to perceive what they're saying. And then somebody, like an Indian speaker, a Hindi speaker, might be uR1ng the rhythm that's, you know, they also know English as a native language, or probably second language, but they speak it really well. But the rhythm is totally different. So for an American English speaker, it's hard to understand, to know where to segment a phrase, to know where one word begins and ends, because they can't, they don't hear the pitch change or the peak of stress to know that, oh, that was a specific word. And like, is that the kind of information you want?

INT: Yes, yeah.

R1: Okay. Um, yeah. So it's definitely, there can be like several sounds, there can be vowels might be pronounced incorrectly. So that, you know, there's common errors that people from different language backgrounds make. So that some vowels are, don't sound like the intended vowel. So you can misunderstand the whole word, because there might be, it might be a minimal pair that, if they use the correct sound, you hear that word, but if they use the incorrect vowel, there's an identical, a nearly identical word that you think, oh it's that word. Um, so yeah. So just distorted vowel sounds, inaccurate consonant sounds. Inaccurate word stress. And not reducing vowels properly, so that every syllable's pretty equal in length. That can make it hard to process and listen.

INT: Yeah. Actually, in my case, the like talking in English with the proper stress or proper intonation is pitiful for now. I mean, really hard. So I try to mimic what the American are saying, but I always put like, you know, the stress in the wrong place. So it's pretty hard for non-native speakers. Korean has no stress or no intonation, so it's like quite hard. So maybe the mastering this segmentary factors are more eaR1er for me.

R1: Yeah. It's eaR1er to teach people how to produce consonants, because you can talk about places of articulation and manner of articulation. It's harder, then, to teach vowels, because it's about how open your mouth is, and the tongue poR1tion.

INT: Yeah.

R1:   But then I agree with you that things like intonation and stress, sometimes it's hard to even perceive the crucial feature, you know, a pitch change or a lengthened vowel can be hard to perceive those cues if your first language handles those things differently.

INT: Yeah.

R1:   And, yeah. I'd say, and in the work I've been doing, phrase stress seems to be one of the harder things for people to modify. They can modify their intonation. They do get so they reduce vowels. But sometimes getting that primary phrase stress in the right place is harder.

INT: Yeah. So like when I tried to order something, like at the Starbucks like that, I think that they don't understand it because my intonation is just so flat. So if I put some stress on several, yeah, major words, then they can understand what I'm saying. So I think that that's the main factor for intelligibility for non-native speakers. So that's my understanding. [laughs] Yeah.

R1:   [laughs]

INT: So like I'm going to ask about, ask questions about the segmentary factors and suprasegmentary factors. First of all, in case the segmentary factors caused the problems in your understanding, what specific phoneme made the listening difficult, like for example, vowel length or individual consonant? Which one is more important for your understanding?

R1:   Mmm hmm. So when you say vowel length, what are you referring to?

INT: Individual consonant, yeah, consonant.

R1:   Oh, consonants?

INT: Yeah, consonant. Yeah.

R1:   So which consonants are more problematic?

INT: Yeah.

R1:   Okay. I would say the "l" and "r," people from language backgrounds where those two are switched. And sometimes it's "l" and "n," but I have more trouble with "l" and "r." Um, I think the "e" [long e] and "i" [short] for vowel—well that's vowels, though. You want to focus on consonants right now?

INT: So which one is, I mean, more problematic?

R1:   Between "l" and "r"?

INT: Between like vowel lengths or consonant or some other factors?

R1:   Oh. I would say often consonants are the bigger problem. Because, I don't really know how to give you a kind of a principled reason for that. It's just more my intuition. Because, for example, if somebody switches the "v" and "w," which is another common one. Or "p" and "f" for Korean speakers . . . .

INT: Yes.

R1:   Can make it very difficult.

INT: Yeah.

R1:   Because sometimes what's being switched is so, seems unusual to an English speaker. So a "p" and "f" switch just seems strange. "L" and "r," even native speakers of English will occaR1onally switch them. I think they have some closer relationship in our brains. But "p" and "f" for a native speaker of English just doesn't, it's just not the kind of mistake a native speaker could make. But even the "t" and theta and "d" and "eh," you know, saying "dis" instead of "this." There are certain dialects in American English where people do that. So I don't think we have as much trouble understanding those kinds of things. But probably some of

the palatals, like the "esh" and the "guh" [soft g] and "juh," when those are not done properly, they might cause some confuR1on. I know the "z" sound for Korean speakers, what's the word. There's an organizational, the "z" sometimes isn't produced right. I know I've worked with students where that, not being able to produce that "z" like design . . . .

INT: Yeah

R1:   Comes out, yeah. And again, it sound, it's two sound differences that don't, aren't, wouldn't commonly be made in English. So, you just wanted to know more vowels versus consonants, which one's more important?

INT: I think that the consonants are more important, but like I don't, Rater 4 said that vowel lengths sometimes like confuse his understanding. So I'm really wondering about that, oh, really.

R1:   Well, when you say vowel length, I think about the duration. But you're not talking about say an "e" [long sound] versus an "i" [short sound].

INT: Yeah.

R1:   Is that what you mean by vowel length?

INT: Yeah. That kind of thing.

R1:   "E" "i"?

INT: Yeah. I think that vowel length is closely related to like stress.

R1:   Yeah, it is.

INT: Yeah, weak vowel or strong vowel, so maybe their point is not noted by Rater 4. I think that's it.

R1:   And I'm sure for different people, you know, certain features will be more important than others. So he and I might have really different experiences with the kinds of non-native speakers we've worked with and how much we've done that. And I've become more conscious of vowels more recently, because I've had some students who have real problems. But I don't think normally I'm focused on the vowels so much as the consonants.

INT: Yeah. Then like from the suprasegmentary factors such as lexical stress, like intonation, or relative duration like that, which one is the most problematic for you?

R1:   Um, what was the last one you said?

INT: The relative duration.

R1:   Relative duration.

INT: Yeah.

R1:   Okay. I think lexical stress is the most important there.

INT: So actually my study is motivated from several misplacement from Phase I. So if you asR1gn, if you asR1gn the test takers below 5, then they go to Phase II, then like they get 5 from the test scores reading aloud like that. Then what do you feel about that or why do you think the reason, what do you think? I mean, why do you think they are exempted from Phase II? So your intonation is different from someone from Phase II. So what's, what could be the reason for that?

R1:   Um, it could either, it could posR1bly be that the topic that was being discussed was either more difficult for the speaker and they had to use some words and sounds that are their biggest weaknesses. It could be that there was just one or two keywords that were missed, and we're instructed that even if there's one word we don't understand, we're to move them on. And so that, if, that's pretty restrictive. So if we go, if we're that, you know, restrictive in not pasR1ng because they missed one, they said one word we couldn't understand, I'd say there's a good chance, or there is some chance that if they took Phase II, they would still pass because maybe

it was just one thing that they said. Maybe they said it very quickly or it was a term that, it could have been a technical term that the rater wasn't familiar with. Like something from their field of study, the name of a town or country or something.

INT: Like, I think that the Phase I, the test from Phase I is more communicative tasks, like interview things. Then Test 2, tasks from Phase II is just reading aloud the script. So . . .

R1:   Oh, I have another answer.

INT: If the student get 5 from Phase II, they can, their performance in the more freer than academic contexts can be, was, than their performance from Phase II. So what do you think about that?

R1:   Yeah. I mean, there's definitely two different styles. So spontaneous speech is one. You're going to get one kind of production, and reading you're going to get a different. My expectation would be that people could do better on the reading than they could do on the first. So that could be why somebody could fail Phase I and come to Phase II and do a better job and pass. Um, yeah. So I think that, that's a good point, that they're two different tasks. And the more important one is Phase I. Because if you can't be understood in spontaneous speech, that's probably more important. Because it's very rare that people read out loud except for maybe a conference presentation or something. But people usually are speaking spontaneously. Yeah. So if they're failing Phase I but pass Phase II, to me it would make sense that they just, they are reading. They have the words in front of them. They're probably being more careful. And it's a different task than just, I have to think what I want to say, and I'm forming my meaning. Now I have to get the sounds to match, and the words to match. And when you're focused on meaning more than form, you're probably going to make some mistakes in form. In the second phase, you're not worried about meaning at all. You're only focused on form.

INT: Mmm hmm. Yeah. Because some TAs said that from the Phase I, they wanted to assign 1 to some test takers. But they didn't do it because some test takers came to the raters then complianes, you know, like, why I am, I got 1 like this. So they just sent them to Phase II. Then that test taker got like 3 or 4 from Phase II. That rater said that it's just so unfair. His pronunciation or his oral ability is totally unintelligible. But he got exempted from the ESL pronunciation class. So she think that that test taker needs the pronunciation classes. So that's the problem. So that's why I ask this kind of question. Then actually, specifically which format do you use in the Phase I? Interview or you just are like seeing what the test takers talking or?

R1:   So how do I do the interview or how do I do it?

INT: Yeah.

R1:   I usually start just by asking them something about, well, how, if it's in the fall, then I kind of assume people just arrived. And so I ask them just to tell me, you know, how long they've been here, where they're from. And from that, I can pick a topic to ask them to talk about for 2 or 3 minutes. So it could be that, you know, they had actually been living in another part of this country, so I might ask them, well, tell me the differences between there and here. Or somebody might have had problems getting here, so I just ask them to explain. Or I might ask someone to describe a place in their country that I should viR1t or a favorite holiday, or even tell me, sometimes I see a major that I don't know something about, and I'll ask

them to describe their major for me. So I think I lost track of your question. Can you repeat your question again? Am I answering it?

INT: Yeah.

R1:   How do I do that first part?

INT: Yeah. Then like, actually, we're employing the group format.

R1:   Oh. I, sometimes I do them in a circle. Sometimes I call people in one by one.

INT: If, in another room?

R1:   They might, yeah. They might be waiting. Sometimes we put in different . Sometimes we're in a big lecture hall, and I sit over here. But the people I'm going to test are over there, and I call them one by one. Sometimes I'm in a classroom, and I have people wait outR1de and come in. Other times, I've had everybody come into the same room, and sometimes they R1t in rows, if the class is set up that way. Sometimes we R1t in a circle. But I just have them go one by one and tell them to introduce themselves. And based on what they say, I may ask another question. But it's very much, I just want them to talk. I don't try to participate, like I'm making conversation.

INT: Yeah. So in this R1tuation, the other test takers are seeing what the other test takers saying.

R1:   Exactly.

INT: Yeah. So I can expect some ordering effect. So the test takers who are interviewed later, they can like, they can know what the procedure is, or what you are asking in advance. So this is going to be beneficial for them, or not?

R1:   Yeah. I think it, people who go first might be more nervous. They don't have preparation time. The people who go later, middle to later, they've had time to think about what they want to say. And I agree with you. They hear what everybody else says, so that gives them some ideas. Well, they can say the same things. I don't, I'd say I've only done that, the group testing, maybe 10% of the time. I really don't like that. I think the only time I've done it is when, you know, we were told to do it that way.

INT: Yeah. But from the related research, the group format, the group testing format is like gaining popularity these days.

R1:   Really?

INT: Because it saves time.

R1:   Uh huh.

INT: And it saves the raters' energy. You don't need to lead any discusR1on or the, by the strict term the group or discusR1on format is just seeing the interaction between test takers. But we don't have any interaction from Phase I, right?

R1:   It's hard to do, because the tester doesn't know their names. So I don't know if I get a discusR1on going, yeah. I can see how that person's doing with that person and that person, but I don't know their names. So in order for me to actually rate them, I have to be able to know, okay, here's your name and you're the one talking. And it's, it would be, unless they all have their names in front of them, and their UINs or something, the best thing would be just a 1, 2, 3, 4, 5, so I can match it up. But we don't do it that way. So, but the advantage of a group like that is people would be more relaxed. And they seem to actually enjoy. The one time I did it, and I don't know why I did it that way, because I had a hard time knowing who was who. But somehow I did. There was some interaction, and people seemed to enjoy the test more. But I think it is very different. It's yet a different way. And some people might be more shy or not comfortable talking about themselves in

front of other people. So that's why I don't really like it as much, because I think it could be very, you could get a misreading of somebody. They might perform more poorly than they really can perform because of nervousness or, you know.

INT: Yeah. Then their oral performance, they are that different, so.

R1: Well, then they'd probably fail Phase I and go to Phase II, and maybe pass Phase II. So it's not a terrible thing. But it puts them on the spot, and maybe is not necessary to make them feel that way.

INT: Yeah. So actually, you do not agree with the group format idea?

R1: Hmmm mmm. No, not really. Not unless I had a way of treating it like a group discussion and a way to identify everybody. But I'd have to have them wearing numbers or something, so I could say, okay, you're 1, and. [laughs]

INT: [laughs] Okay. I see. And then, like so I wonder if from the training sesR1on before the EPT, you got some evaluation criteria from the training sesR1on?

R1: For doing Phase I and 2 together. But we never, I have never been trained on how to do Phase I.

INT: Yeah.

R1: So sometimes we do them together, so we go into the room, and we're doing one person at a time. And we decide. Do we only do Phase I, or do we go ahead and do Phase II right there? That's sort of the most common way. But when we do that early morning evaluation where it's Phase I only, I never received any training on doing that. But we're told, I mean, we're told to use the same criteria that we do if we're doing Phase I and 2 together. So I do that. It's if you can, if you have trouble understanding even just a few words, and it's not, you just try to get them to talk as much as posR1ble so you can get a good speech sample. But in terms of doing, how to do it as a group or, you know, just more the practicality of doing it, I never received training. So I always felt like I was sort of winging it, you know, kind of guesR1ng how to do it. But I just did it the same way as when I do the full interview, or Phase I and 2 together, where you choose, maybe you exempt someone, and they don't finish the test. Or you choose not to exempt them, and they go ahead.

INT: So like I heard from the EPT faculty advisor that there's just so much uhm…about the ESL course, pronunciation course. So do you think that we need more measures in communicative abilities from the oral section for EPT? Or is it just okay if we measure only some pronunciation or intelligibility?

R1: Hmm. Well, I mean, it just, you have to decide what, I mean, the purpose of the test, my understanding is, that we're screening people for pronunciation, and then we have these classes, you know, 510 and 504 that they can take to help with that. We don't offer classes that help people with communication. [laughs]

INT: Communication [laughs].

R1: Yeah. We have writing classes, or we have the ITA classes, but we don't have a specific class on oral communication. Maybe one of the buR1ness, the buR1ness class might cover a little bit of oral communication. So the problem is, if we're testing and rating them on their ability to communicate, but we don't have any resources to help them.

INT: Yeah. So I mean like do you think that we need more like ESL courses for communicative ability?

R1: Yes. If you're talking about the courses, yes. Yeah. Often what happens is students, they do benefit from the pronunciation classes, and they do get some help in, especially 504 and 506, they learn more about communication. 510 there isn't

much opportunity for communication. It's all about pronunciation. But often students still feel they have trouble knowing how to communicate effectively.

INT: I need it. I need it. [laughs]

R1: Yeah, and everybody feels that way to some extent, you know. And a lot of students, as you know, maybe they interact in their first language community while they're here. And so I work with a lot of students who, like their lab is all Chinese students, and they're a Chinese speaker. And their advisor is a Chinese speaker. So they can live day to day just speaking Chinese. And then they have to go somewhere and use English. And, you know, the communicative ability, the socio-linguistic ability, they're just, you know, they feel less comfortable. So certainly, I would love for there to be other classes that help with conversation, you know, oral communication skills. Academic skills, maybe a little bit of everyday English type stuff.

INT: One month ago I went to a language testing research colloquium to Cambridge with The EPT faculty advisor, and with like some of the doctors from Education Psychology. Then I'm very interested in oral section and oral testing, and the major concern was that, like about interactional competence. So how do we measure the interactional competence? And what the interactional competence like the construct, what interactional competence, so that's why I'm, I got interested in the more interaction, interaction things, the more communicative things we have to employ from our, to our ESL courses. So that's the question I ask to The EPT faculty advisor that, yeah, there's a lot of debates, but we don't have any courses like that.

R1: Yeah. Some common difficulties I have is sometimes I don't know that someone's asking me a question because of the way they phrased it. And it's not word ordered. Like it's not, I, every time I hear it, I wish I would write it down, because it's things like I want you to tell me this, or it's stated as a, it's phrased as a statement, but not as a question. But it's also done in a way that I feel like they're telling me something instead of asking me. So to me that's kind of an interactional or, you know, form-related competence that, it's very interesting. And then there's also, as you know, when we talked about this in socio linsuistic classes, it's just that a non-native speaker can come across as being too direct or demanding, when they don't mean to be, but, because they don't realize how to be appropriately polite and how formal they should or shouldn't be. So what we do in the SEC. And, yeah. There's, those are definitely difficult things to learn in a second language. So . . . .

INT: Yeah. That will be all. Thank you so much.

R1: Okay.

INT: Yeah.

R1: Let me get this turned off.

(end of interview)

INT: It's working.

R2:   Yay! [laughs]

INT: Actually, I need to get some information about yourself.

R2:   Mmm hmm.

INT: Then, so you are a native speaker, right?

R2:   Yeah.

INT: And then, how many years have you taught ESL classes?

R2:   Mmm, five, five or six, yeah.

INT: Five year, yeah. Did you teach like pronunciation class or writing classes?

R2:   Pronunciation, the whole time. Yeah.

INT: The whole time, yeah. And then ESL 510?

R2:   I had one semester of 510, and I've taught approximately 10 semesters of 504, because they offer it in the summer, too.

INT: Yeah.

R2:   Maybe one semester of 506. And, uh, I mean, I've taught some other classes for the IAI and stuff, but mostly it's 504.

INT: 504.

R2:   Yeah.

INT: Can you remember how many times like did you rate the test takers from EPT?

R2:   From EPT?

INT: Yeah, or a section.

R2:   I think, I truly think it's every semester since I started teaching, so that would be approximately 15. Because fall, spring, summer.

INT: Yeah.

R2:   So there's maybe one or two that's missing, but between 13 and 15 times.

INT: Yeah. So the, how about the whole English education as a second language, maybe how many years have you taught second language learners English? The general English.

R2:   Oh, it's the same number, so 5 years, mmm hmm.

INT: Five years. Then you are in the department of Linguistics?

R2:   Mmm hmm.

INT: Did you attend to any single program, Applied Linguistics?

R2:   No, no. I did, I got my MATESL here in 2004 to 2006. And then, no, 2004 to 2007. And then 2007 to now, I've been in the Linguistics department. So it's, but I haven't ever been, I'd never took any other college classes, other than here.

INT: Yeah. So when you were undergraduate student, what was your major?

R2:   My major was Spanish, and my minor was ESL.

INT: So you have.

R2:   Yeah. [laughs]

INT: Okay. Thank you. What do you think is the purpose of Phase I of the oral section, EPT oral section?

R2:   It's to, so the main test is to put people into 510 or not, and sometimes that's a difficult decision. But sometimes with people it's really easy. It's like clearly you do not belong in 510. So it's kind of to move those people through quickly so that you don't have to take time. So to identify people who are way above or way below the middle.

INT: Yeah. So like in rating Phase I and Phase II, both of them.

R2:   Yeah.

INT: And then do you think they are the separate measure, or just Phase I oral test is some complementary device for Phase II? Which one?

R2:   Hmm. Well, what do you mean that they work together?

INT: I mean that Phase I works, worked independently as a measurement, or just complementary for Phase II? Which one is true for you?

R2:   Hmm? I mean, so the two things, the Phase I and the reading of Phase II, seem to me to operate separately. However, they're, they work together in the sense that Phase I gets people out who don't need to take Phase II. I don't know if I'm answering your question right, because I don't know if I understand.

INT: Oh yeah, that's okay.

R2:   Okay.

INT: Yeah. So which factors of the test takers' performance do you focus more, in the Phase I?

R2:   So I wholly focus on pronunciation. And within that, whether or not there are ever times when I don't understand a word. So I don't care at all about their grammar, their fluency. I don't care at all.

INT: Oh, really?

R2:   Yeah. It's only pronunciation, and it's also, do I ever have to stop and go, what was that word? So if I have no problems in communication in terms of understanding the word, they move on. But if there's even just a couple words that I had to take extra time to think about, then I move them through to the second part.

INT: Oh. So what made you decide that the test takers wouldn't be exempted from the oral section Phase II?

R2:   Yeah. So if they ever have words or phrases that I don't understand, they have to do Phase II.

INT: Yeah. So when do you assign 5 to the test takers? When you understand every word they uttered?

R2:   Yeah.

INT: Okay.

R2:   And that doesn't mean that every word that they uttered was perfect. It just means that, I mean, they have an accent, probably, but it just means that I didn't have trouble understanding the word. It was very easy and effortless for me to understand every word.

INT: Oh, yeah. I see. So, you mean that like their pronunciation is not perfect?

R2:   Right.

INT: Not there like, but still you can understand what they are saying?

R2:   Yeah. Yeah. I mean, because the people who get 5s, it's not as if they all sound American. It just means that their pronunciation in no way hinders communication.

INT: Uh huh. So I'm thinking about this kind of a case. When the pronunciation is easy to understand, but their grammar is very bad. And then you cannot understand what they are saying.

R2:   Um, but, so, I would differentiate between understanding like a general message, and understanding every word.

INT: Yeah.

R2:   So it's possible that I'm having difficulty comprehending the main point of what they're saying, but I can understand all of the individual words. That person would, for me, would pass the EPT oral. Because it's, for me, the purpose of that class is, do they belong in 510 or not. And 510 is, does no work with grammar, anything.

INT: Yeah, yeah. I have taken 504, so I know a lot of this.

R2:   Mmm hmmm, mmm hmmm.

INT: And, so when you assess the test takers' oral performances, do you focus on the recognition of a word form or like the overall construction of meaning?

R2:   Recognition. What do you mean by recognition of word forms?

INT: I mean, that you focus more on phonemes?

R2:   Yeah, I would say that's probably true. I mean, I'm trying to understand what they mean. But it's more important that I can understand all of the words that they're saying.

INT: So more specifically, what issues prevent you from understanding what test takers are saying or made the test takers' oral performances unintelligible?

R2:   Well, I mean, that's, for me that's totally related to the individuals. I mean, sometimes people don't have any kind of rhythm or intonation that's native like, and so I'll have trouble understanding a word. Maybe they put stress in the wrong place. Maybe they don't put word stress. It's possible that they have sound level issues. Like they have trouble with, you know, "l" and "r," or "p" and "f," and they make mistakes that make the word unrecognizable to me. So it's, I think a lot of different factors that are really just related to the individual.

INT: So will you tell me some specific examples of a problematic word or phrases or sentences the test takers saw.

R2:   Um, it's actually been a whole semester since I did this.

INT: Yeah, yeah. I know.

R2:   So I can't think of any specific examples. But you know, I'm giving the EPT next week. So if you want to ask me that question again, after, then . . .

INT: [laughs] Yeah.

R2:   Really fresh, and I can probably give you a specific example. [laughs]

INT: Yes. In case that the segmentary factors cause the problem in your understanding, what specific phoneme made understanding difficult in your cases?

R2:   Yeah. So again, I can't think of a specific time. But I mean, common problems that often cause problems with intelligibility, "r" and "l," "p" and "f," sometimes "l" and "n." Could be "z" "do" like "juh." I mean, all kinds. That's probably the most common.

INT: Most common. Okay.

R2:   Maybe like the "th" and "the," like theta and "ev" is possible.

INT: "Ev," yeah.

R2:   Those don't, I mean, those, I don't think, are as big. But they can cause problems.

INT: Yeah. So how about vowel lengths?

R2:   Hmm, yeah. Um, oftentimes "ee" and "eh", "a" and "ae," and "uh" and "ah" and "a." Like those, so those pairs. Those three pairs.

INT: So do you think that the vowel length, does it make problem like individual consonant?

R2:   Are you asking me if I think vowels are less problematic than consonants?

INT: Yeah.

R2:   I don't know. I never thought about it. Um, maybe. Maybe, I would say that that's possible. But I, I couldn't tell for sure. [laughs]

INT: [laughs] Yeah. So in case that the suprasegmentary factors cause problems in your understanding maybe, what specific factors would be more problematic?

R2:   Definitely word stress.

INT: So you mean lexical stress?

R2:   Yeah. Lexical stress. You know, if someone says "obligatory" [long "a" sound] instead of "obligatory," I mean I know that example, but if they say it on a word that I'm not used to hearing it, sometimes I just have no idea. But I think also just general phrase rhythm. I mean, if the person's not reducing function words and stressing content words, oftentimes that can just get difficult to listen to.

INT: Yeah. I see. That's my problem, I know. [laughs] So my study is originally motivated from some misplacement issue. So some raters thought that when they assign a 4 or a 3, then they go to Phase II. Then they get 5 from the Phase II. But they don't agree with that opinion, because their pronunciation is not perfect. So in that case, what would be the problem? What would be the cause for that kind of misplacement?

R2:   Mmm hmm. Well, I think that, from my understanding of how the test works, Phase I is quite strict. So it has a really strict criterion, and that way it doesn't allow too, it doesn't allow people who should be tested to go through. So it's very possible that like, you know, you're going to test somebody, and they're going to be an exempt. But the one thing I don't quite understand about your question is, for me, like I don't think, when I interview a person in Phase I, I don't think, "oh, they're probably a 4 or a 3 or a 2." I think, are they a 5 or a 1? And if not, I test them. And so, okay. So then if the person ends up getting exempt, but they don't have a perfect pronunciation, the other thing is, is on Phase II, you can get answers wrong and still get exempt. Because there's like 150 points or something. And I, or I don't know how many there are. But, so it's possible that the person does have noticeable problems, but just not enough that, not enough and not of the type that they are going to need to be in 110, 510. You know, there's a lot of, so let's say the person has a lot of trouble with like one sound group, like "l" and "r." They mess up "l" and "r" all the time. Well, on the actual test, there's only 5 points related to "l" and "r." So maybe that person doesn't get those really that wrong, and is, does really good with everything else. Well, of course they're going to be exempt, because they have one problem, and maybe it's a major problem. But each little thing is only worth a couple of points. So that could happen. I don't know. That's never, I've never had that experience, where I thought, "wow, I should test this person." And then they got zero wrong or something like that. I've never had that experience. So it also could be maybe that they're not good at doing the first part. [laughs]

INT: [laughs] Yeah. From my experience, like I took EPT. And then maybe, finally I got 4 from Phase II. But I didn't know that the Phase I is about pronunciation. And then it's very hard to focus on pronunciation. Because I think that it's more like the communicative activity.

R2:   Yes.

INT: Yeah. So, so the raters all, might think that my pronunciation is problematic from several perspectives. Then she or he like sent me to Phase II. Then I got 4 from Dr. Dickerson. But I think that I need more, then I'm exempted from the ESL 504 or 506. But I think that I need this. So I took, kind of, that classes. So Phase II is like easier than Phase I, I think. And at that time, actually I got, a linguistic class from English department in Korea. So I had some knowledge about pronunciation or rhythm or stress like that. So I think that I could do better, reading . . .

R2:   Yeah, do the reading.

INT: My oral ability, like that. So . . .

R2: But I think for me, the way I would react to that is to say, well, we want the Phase I task to be the most difficult. And we want the Phase II task to be related to the types of activity in practice that you will have in the class. And if you can be successful at the types of activities that they do in the class, then you very well might need practice to transfer that skill to your every day talk. But you don't necessarily need the class, because you can do those things. And the people who belong in the class are the people who, even when they try, they can't do those things. Yeah. So that's what I would say about that.

INT: Yeah. So Sun Ju gave an example. She gave the, in the oral Section 1 of Phase I, she wanted to give the score 1 to some student. And his pronunciation is totally unintelligible, she thinks. But she just sent him to Phase II, and he got like 3 from the test. So she think that it's unfair. He needed some pronunciation class, but he, he was exempted from taking classes. So what do you think about this case?

R2: Um, again, I've never had, I don't think I've ever. I don't think in my life I've ever given a 1. Maybe one time. That's pretty rare for me. So 3 means that the person has quite a few difficulties, and it means that they're, they're strongly recommended to take the class. But I think it also means that, I mean, the, in my understanding the class 510 is designed for a very small percentage of the population that comes here, that just struggles with communicating in English. To the point where, like I had, I think I had a student, maybe just one, like a student who ended up going home because they just could not communicate with people in the U.S. So to me, that's like such a far extreme. And it's not because of their grammar or their fluency. Like literally, the sounds that come out of their mouth make it impossible for them to talk to other people. And I guess that I would say that that population is probably quite small. So even if you have trouble understanding someone, I mean, I think to get a 1 on Phase II is crazy. I don't know. And so, I don't know. Maybe it's possible that that happens. But I don't think it probably happens very often. And I think it's also important to remember what the class is for. I mean, the class isn't for like teaching assistants, or something. The class is like, are you going to make it? And what a 3 means to me is, you're probably going to make it, but it's going to be hard. And a 1 to me means, wow, like you might not make it. And so you better try really hard in this class.

INT: [laughs] Yeah.

R2: So I don't know. I don't know. And I guess one other thing I would say is, every once and again, as like a pronunciation teacher, I run into people that, I mean, I've been teaching pronunciation for so long, I've been communicating with people for so long that have troubles speaking English. So I feel like generally I'm pretty good at it. But every once and again, I meet somebody that I just seem to have a really hard time with. And I don't know if it's something unique about that individual. So it's possible that Sun Ju just talked to somebody that was very different from her experience.

INT: Yeah. Okay. [laugh]

R2: I don't know. Yeah.

INT: And like, did you get any training session before the rating, the oral section?

R2: The EPT?

INT: Yeah.

R2: Yeah, definitely. In fact, still do every semester except for the summer. There's always like a one hour. So I took EIL 488, and there's like a section in that class

about it. And then every semester before we give the test, we go over the purpose. We watch a video. We do practice. We get feedback.

INT: Yeah. Then, so, which factors from the evaluation criteria do you rely more?

R2: For which part?

INT: The, you've got some evaluation criteria from Dr. Dickerson. And then which one do you think is the most important?

R2: Most important?

INT: Yeah.

R2: Well, really honestly, for me, the evaluation criteria is the evaluation criterion that he gives us, which is for Phase I. You have to understand every word. If you don't, send them through. And so my thing, that's what I, that's how I rate it, is if I understood every word or not. And then for the second part, the criteria are just related to whatever feature. So, does that answer your question?

INT: Yeah, yeah. Everybody said the same thing.

R2: [laughs] Yeah. It's pretty simple.

INT: [laughs] And then, actually, the, my another point is about the format. And then we seem to employ some group format in the Phase I. How do you assess the people you are interviewing, one by one or you are just observing what they are saying in the group?

R2: No, I definitely do it one by one. For me, I think the only reason that doing the Phase I as a group saves time is I don't have to introduce myself to everyone who walks in a room, right? But the way that I do Phase I is literally, everybody is sitting in a little row. I tell them, say your name, say where you're from, talk to me for 90 seconds about why you're here, and something interesting about you. What you did this morning, I mean, whatever, just to get them to talk. But it's completely individual. They're talking to me, and I just go, boom, boom, boom, down the line.

INT: Yeah. Then the other test takers are watching . . .

R2: They're listening, mmm hmm.

INT: What you are, yeah. And is there an effect because they were watching what other people saying, and then watch the procedure and the, some expectation of what you will ask the questions, the prior knowledge about the interview questions. So is there any effect on score like that?

R2: Possibly. Although, I mean, it's not as if they get to see the procedure for the person before them and then they get this, a similar question. I mean, literally when I do the group thing, I tell people. Like I just want you to talk at me for 90 seconds. I don't care what you say, just say something. So they definitely get to see the procedure, the exact procedure that they would get. And, I mean, for me, that is one of the things that Dickerson always tells us to do is ask different questions, so that people can't prepare. But I think that, I believe that because I don't care about fluency or grammar or vocabulary, I only care about their pronunciation.

INT: Oh, yeah.

R2: And if they have trouble with word stress and rhythm, they're going to have trouble with it whether they have 20 seconds to prepare or not, you know.

INT: Yeah. I know this. So the intonation or pronunciation cannot be mimicked, or.

R2: Yeah. Or it can't be improved upon in 3 minutes while you listen to other people.

INT: Okay. I see.

R2: I mean, that's how I feel.

INT: Yeah, yeah. That's the answer, yeah, I think. Yeah, so . . .

R2: Certainly their grammar and stuff might be better, but, yeah.

INT: Yeah. So did you allow the students interact each other in English?

R2: In the Phase I group?

INT: Yeah.

R2: No. I think one time a couple years ago I tried it. But I just found that it took longer and was . . .

INT: You cannot assess what you're able to measure.

R2: Yeah. And it just seemed irrelevant. It seemed more time consuming. So, I mean, I'd be open to it. But for me the whole point of doing Phase I as a group is to be faster and more efficient. So I don't know why they would need to talk to each other to do that. And I think it would be awkward for me to go and sit and listen in on two people talking. I think they, that would just make them feel uncomfortable, so.

INT: Oh. So if we, we allow the test takers interact each other, then you're not going to participate in that discussion, then you're just watching what they are saying. That's called a group discussion, or group discussion format. And that issue is quite popular these days. And that's the other point about my thesis. And the advantage is that we can save time and energy like that.

R2: I guess my question is, how would you get them to talk to each other, and how would you ensure that just a few people didn't control the talk and other people never said anything.

INT: Yeah. It's true. So the, they have the different proficiencies, the different characters. So it's hard to control, but they still. Some studies reported that like they feel more comfortable when they talk to each other. Because they don't need to talk to some interviewer.

R2: Yeah, yeah.

INT: Who is more, you know, native speaker, so.

R2: Yeah. Mmm hmm, mmm hmm. Yeah. I mean, I think that that sounds like it has good advantages, but I personally just never felt like taking the time to deal with the disadvantages. And the other thing is, is it goes very quickly. So I would say that when I give a Phase I group, it never lasts longer than 20 minutes. It's usually about 10. So just expecting people to start talking to each other, while they may feel more comfortable, I think, I think it would take more time. But I think if you organized it in a certain way, it might work.

INT: Yeah.

R2: And that would be nice, because they I wouldn't have to listen to everybody. But I don't know.

INT: And then the last question, do you think we need more, like some major to assess the communicability rather than just pronunciation? And then like, do we need more ESL courses that is supposed to teach some communicative oral ability?

R2: Uh, do you mean do I think there is a lack of oral communication classes that are unrelated to pronunciation?

INT: Yeah.

R2: Definitely. I think that is the one thing that, there are none. There are no classes offered by the University to support talking that is unrelated to pronunciation. And all the time my students ask me for it, and all the time. I mean, there are probably things in the community. But the University doesn't have them. And I really wish that they did, because I think a lot, I know a lot of the problem is people just don't

have an opportunity to speak English. But on the other hand, if you have a class full of international students getting to talk to each other, that's great. But I can understand why the University wouldn't want to offer that, because they'd probably say, hello, like they live in the U.S. You know, if they want opportunities to practice and to speak and to communicate, they should go out and make friends and join activities and do those things. So part of me really wishes that we had those classes. Because I think a lot of people need them. And they don't need pronunciation. They need to learn how to interact with people better. But another part of me says, those same people would probably benefit just by trying to use English more. I don't know.

INT: So I think that we need more like to improve the oral ability related to academic contexts, like presentation or talking to professor and like discussion like that. So I think the, in my case, I need more academically related communicative oral ability, just for the pronunciation. So that's . . .

R2:   Yeah. And 506 is kind of similar to that. But it's really focused to teaching. So if you are just trying to be a student, it's a little bit different, you know. And certainly there are a lot of international students on campus that want to be successful students. But they don't care about teaching ever. You know what I mean? Like they, they just want to have a support system so that they can perform better in their classes. Yeah. I mean, I think that the, I'm always surprised by that. Because we have pronunciation and we have writing. But we don't have any academic oral English class. We don't. And like IEI does. They offer a class like that, were they do presentations and group work. But we don't have anything like that. And I never understood why. Although I think, I think like ESL 505, it's something. I don't know. I feel like it's something to do with the business school, might have like an oral component. But I think you have to be like a business student to take it. I don't know.

INT: Yeah. It's very, like it's just for a business purpose like that.

R2:   I think so, I think so. I'm not sure.

INT: Yeah.

R2:   But yeah, that'd be great.

INT: So the EPT faculty advisor said that there's much debate about the introducing new courses about communicative oral ability. But that's not going to be easy to, yeah, set up some new service courses. So . . . .

R2:   Yeah. Yeah.

INT: It's a lot of work, lot of money, like that.

R2:   Mmm hmm. It's probably very expensive.

INT: That's all.

R2:   Okay, good.

INT: Thank you so much.

R2:   No problem. Hopefully you got all your answers . . .

(end of interview)

Rater 3

INT: So what was your major in undergrad?
R3:   Undergrad, my major was graphic design.
INT: Then how about the grad school?
R3:   Grad school master's was TESOL, and now Ph.D. it's Curriculum and Instruction.
INT: So what's your interest of research?
R3:   I'm interested in Japanese learners of English and their oral production of English.
INT: So how many years have you taught the second English learners?
R3:   Total probably about 10, 13 years, I believe.
INT: Thirteen years.
R3:   Yeah.
INT: So do you have any periods to teach pronunciation?
R3:   Yeah.
INT: Yeah. So how many years?
R3:   The past 4 years.
JEL   Four years. So ESL courses?
R3:   ESL courses, yeah.
INT: ESL 50 . . . ?
R3:   504 and 506.
INT: So how many years have you experienced rating EPT?
R3:   Uh, let's see. I think it's been the past 4 years that I've done it. Yeah. So I think every semester for the past 4 years.
INT: So have you taught English at other than in America, like other countries?
R3:   Yeah. In Japan and in Korea.
INT: Where in Korea?
R3:   At Chichi do.
INT: At Chichi do. Oh. So do you think you're familiar with Korean accent or a Japanese accent rather than other, Indian or?
R3:   Yeah. I'd say, I mean, my, some accents I'm not so familiar with. So yeah, there is a difference, I believe.
INT: Yeah.
R3:   [laughs] Okay. Repetition there.
INT: The, what do you think is the purpose of a Phase I of the EPT of a section?
R3:   Well, I think it might be the first phase is just to weed out the people who really don't need to move on to, or to have specific pronunciation instruction through the university. I see that as just a way of kind of delineating people who really need instruction from those who more than likely don't need any.
INT: Yeah. So what made you decide that the test taker wouldn't be exempted from the Phase II or not?
R3:   Well, I kind of go on the guideline provided by The test supervisor that, you know, if, are you understanding everything the person is saying. And if there are words that I'm not catching, if my listener effort is high, then I think, well, maybe that person should go on to the second stage. So my main criteria is am I understanding everything that the person is telling me or not.
INT: So I'd like to know specifically which factors of the test takers will have preferences. Maybe you feel like they're intelligible or not.
R3:   Right. Again, a lot of it has to do with maybe segmentals. So consonants and vowels that they're really not producing accurately, and so that's causing me to

guess at what word they're trying to use. Intonation, maybe not so much. To me, in my opinion, it's more of segmentals. And also word stress. If they're really inaccurate with their word stress that causes me to kind of guess and listener effort. At the level of rhythm, again, I'm going out on the guideline of am I understanding every word that they're, whether or not their rhythm is good, uh, that's kind of another issue. But am I understanding the words that are coming out?

INT: So the, which factors, which specific factors, may be you most hard to understand, segmentary mistakes or suprasegmentals?

R3:   Yeah. Again, segmental and word stress. I mean, if they're, you know, there are instances where the word stress is in an inappropriate place, and I really can't catch that word, and so that can really cause a difficulty.

INT: Yeah. So that's the most problematic thing in your understand?

R3:   Most problematic would, I don't know, again, segmentals and word stress. I think it's a combination.

INT: Yeah. So then, so can you remember and discuss which specific factors make you annoying or make you . . . ?

R3:   Annoy me? Um, again, it depends on the language group. I remember having some Spanish speakers, and they're, some of their segmental difficulties were problematic with the, for example, the "ja" "ya" distinction. I don't know if it annoys me. I wouldn't say it annoys me. But it's just something that's kind of problematic, because it's frequent. With maybe Korean speakers, with the final "g" at the end, like language "g", change "g." So those are kind of clear things. With Japanese speakers, it might be instead of saying for, saying hor, using an "h" sound instead of an "f" sound. So I don't know. To me it's language specific.

INT: How about vowels?

R3:   Vowels? I'm trying to think. Yeah. I mean, when I think of vowels, I often think of Chinese speakers having difficulty, like for example with the vowel "a" or "eh," distinguishing between those. Maybe "e" "i" distinction; "ou," having difficulty with "ou." So when I think about vowels, I know Chinese speakers, from my experience of learning, that they have a lot of difficulties.

INT: Yeah. So consonants are, cause more problem than really than vowels do?

R3:   Uh . . .

INT: Not consonants?

R3:   I couldn't rank it. Um, I don't think I can rank it. But, you know, some language groups, the "v w" distinction causes some problems. I don't know if I, yeah. I think it's kind of language group specific.

INT: So, I'm very curious about your concept of intelligibility.

R3:   Okay.

INT: If some letters are very ungrammatical . . . .

R3:   Right.

INT: Then, but you can understand every word right away, sentence they utter.

R3:   Right.

INT: You think that they're intelligible or not?

R3:   Well, in terms of the EP and EPT . . . .

INT: EPT, yeah.

R3:   You know, we're just, as far as I know, the guidelines are specific to pronunciation. So their grammar could be terrible, but if I'm understanding the words that are coming out, it might make a difference in how I rate them. You know, intelligibility, yeah. I mean, grammar is tied into that. So if they're really having trouble at the

grammatical level, I don't know how much of that is related to pronunciation, or I don't know how much pronunciation instruction would help them with these grammatical issues. And, yeah. I think it effects intelligibility, definitely, if someone is speaking ungrammatical. Listener effort is high. You may understand all of the words that are coming out, but in terms of how they relate to each other, you have to ask questions to try to find out the relationship between ideas or words or subjects and verbs. So definitely it's hard. [pause] Maybe there should be an aspect of the EPT related to specifically to grammaticality.

INT: So I'd like to ask about your, which format do you employ in EPT? Some raters take into everything, or other raters take like interview formats. So, yeah. Which one do you use or?

R3: I've used both.

INT: Both?

R3: Both versions, and again, it just depends on the logistics of how many people we have and, you know, typically I use, if I'm given a large group of, you know, 20, 30 people, I've seen somebody take them individually. I tend to just take them as a group and have them individually introduce themselves or ask them questions, follow up questions. So both formats. In terms of preference, I would prefer individual. I like meeting people one-on-one and probing them kind of individually. And I think there's also a social factor involved when you're standing up in front of a group of people. I think that kind of effects how you talk and maybe the accuracy of your production.

INT: So you did not allow the test takers interview each other?

R3: Yeah. There have been occasions where I've encouraged questions from the audience. But typically they don't want to ask questions, or they don't have questions to ask, so I'm the one to do that, yeah.

INT: Yeah. So when you have to take a good discussion format, how do you think about it? Like is it, does it provide them the like appropriate context to assess their intelligibility or not?

R3: Uh, so does a group format provide a good context to? Yeah. I mean, again, I think there's a social factor involved in that standing in front of a group of people is maybe different. But again, there are instances in their, well, there are instances where they will have to do that. You know, here at the university they'll have to get up and give a group presentation, or a presentation to a group. In terms of intelligibility, possibly, you know, distance from me, and if they're at the back of the class, that might affect intelligibility, if they have a soft voice to begin with, it might affect intelligibility. Whereas if it's one-on-one conference, you know, I'm there with them. I can lean in. So yeah, I think it effects to some degree whether or not I can catch everything that they are saying.

INT: Mmm hmm. And when you take the one-on-one interview or good description format, how did the test takers react differently? Did they behave differently or not?

R3: It seems that some test takers, when it's the group format, they've introduced themselves before for other, for their own departmental type orientations. So they kind of have a set format of how they introduce themselves. Whereas on the one-on-one, I don't really, I have specific questions there where I don't let them kind of talk, introduce themselves on their own. I have a question that I ask that may not be necessarily a self-introduction question. Does that make sense? So in this kind of group setting, maybe they can fall back on this kind of self-introduction

that they've memorized to some degree, or if they've given it before, possibly. But then I always ask follow up questions that, you know, probably.

INT: Okay. So, in which situation do you think the test takers feel more nervous?

R3: Um, yeah. I mean, overall I would say in the group situation. I'm guessing. I'm trying to think of instances where someone was noticeably nervous in the one-to-one. Maybe not as much. Again, with the one-on-one, I'm thinking. With the one-on-one, it's typically in the afternoon when they've gone through all the testing already, and they're kind of tired, and it's just. But a lot of the group interviews are in the morning, first thing. It may be kind of, just, they haven't warmed up yet maybe. And so that might make a difference, like time of day might make a difference in terms of nervousness. And also the format of being in a group.

INT: Okay. Actually the whole population who has to take the EPT is increasing recently, because the university is accepting more international students. So they have issues about that, that we have to employ the group discussion format more when the raters don't like it. So do you have any suppositions about employing this kind of a format more accurately or more properly?

R3: Yeah. Um, I think I remember in the past, this was like 8 years ago. I think the class sizes were smaller, I want to say, in terms of the group discussion. Maybe, and if I remember right, we broke, there were two people in one group. So for example, if we had a class of 20 or 30, we had two groups, two circles. And at that time, like more discussion could be between students and things. Um, you know, I prefer one-on-one. I mean, for the first stage, for the first stage, I think you can get a good, fairly decent sense of a person's speaking ability through group discussion, through group format. If there are any questions or if there are any concerns about, for example, if a student, if you're having difficulty hearing a student in the back, then it's always safe to say, well, let me just let that person go on to the second, second rating. You know, it does take a lot of time to do it individually. And it just might not be feasible. And I don't know if there is a happy medium between those two. Um, you know, maybe smaller groups, 10 people rather than 20 or 30, where, you know, picking up the pace a little bit. I don't know. Yeah.

INT: Okay.

R3: It's a tough question. It's a tough question, you know, how to maximize the resources with what you have.

INT: [laughs] Yeah. Mmm hmm. More raters.

R3: Yeah. More raters, and.

INT: Have you have a superior saying misplacement?

R3: Misplacement?

INT: I mean, if you assign the low score . . . .

R3: Right.

INT: In the Phase I but that student can be exempted from the Phase II. So what might be the cause about that?

R3: Yeah. Again, maybe you know, in that group format, maybe again the person might just have a soft voice and you're not able to hear everything that they're saying. You kind of have to guess for some things, so you say, well, let me move them on to the next level. It seems like, you know, these days there are very few who move on to the second level. I mean, the majority have passed that . . . .

INT: exempted?

R3:   Yeah. Exempted that through the first meeting. I don't know if it's because of the quality of the students is improving or what. But it just seems like more are able to move through. And when I'm doing the group interviews, it seems like I'm understanding what they're saying. And they're, but in terms of misplacement, you know, I guess it's always better to err on the side of, to have them go through the second test, even if they fly through it with no problems, or very few problems. It's better just to go through that stage. I think it's worse on the other end if, you know, you're, if they're just passing through and they really have some serious problems, then.

INT: So do you think that, like the different raters or like different tests might be the cause for misplacement?

R3:   Yeah. Possibly. People are more, like for example, like I said, I'm more familiar with Korean maybe, or Japanese accent. Maybe less familiar with Indian accent or Eastern European. So my unfamiliarity might cause me to have those people pass through to the second rating. I'd say that's possible, yeah..

INT: Then . . . .

R3:   So maybe, yeah. Yeah.

INT: Task is also different, because . . . .

R3:   Task, yeah. I mean, you know, depending on the questions that you're asking, what kind of follow up questions. Are you just having them introduce themselves and that's it, or, you know.

INT: Yeah. So do you have, do you think that it's like about the format or about the rating?

R3:   I think it's pretty, I mean, it, to me, is, it seems like the standardization of the individual meeting has been done pretty well. In terms of the group meeting, maybe less so. I think this is kind of a newer format, and in terms of having papers or materials to give to graders to tell them what to do, it seems like there is a need, in my opinion. You know, I mean, the calibration sessions that we go through are calibrations showing one-to-one interviews, not group interviews. So sure, I think having a calibration session where you discuss group interview techniques or problems that you may encounter, or again, having maybe a written format, written criteria. I think that would be beneficial.

INT: Oh, yeah. So actually the, actually we're assessing the intelligibility in the whole EPT oral sections. Do you think TOEFL or other speaking assessment tools are really more like the different assessment, it measured other oral skills rather than just intelligibility?

R3:   Well, it seems like right now, I mean, the test is really focusing on pronunciation, not on intelligibility. And intelligibility, like you said, it includes that grammatical aspect, which we're really not supposed to be listening for. So, I mean, if that's an important consideration, if intelligibility is viewed as being important for the program, then, yeah, I mean, it seems like a separate component would need to be developed, or a different phase, like a third phase, would need to be developed, possibly.

INT: Yeah. So we need to , we need…

R3:   But it seems like, yeah.

INT: Other ESL courses?

R3:   It seems like in the oral courses, I mean we do touch on problems with intelligibility related to grammar. But we don't focus on it. I mean, we, we don't, as far as I know, we don't have materials specifying grammar points, or we're not

fixing on grammar points. So if grammar is a serious concern in its own right, separate from pronunciation, or in conjunction with pronunciation, then I think as maybe a separate course could be developed. You know, it seems like students, some students maybe want a grammar course. You know, and they could take the writing course, which maybe offers some grammar feedback. But there maybe be a different between oral grammar and written grammar, that maybe they're just not getting feedback in the type of grammar that they, in terms of oral grammar. I don't know if that's too much of an answer there.

INT: Maybe that would be all.

R3:   Okay.

INT: Yeah. Sometimes after interviewing, I just come up with maybe some more questions, so . . . .

R3:   Sure, sure.

INT: If I have further questions, I'm going to email you.

R3:   Yeah, definitely, yeah.

INT: Thank you so much.

(end of interview)

What's the purpose of phase I of the EPT oral section?
To weed out the highest and lower performers, to cut down the numbers of subjects that will require the lenthier test.

What made you decide that the test taker wouldn't be exempted from the oral section phase II?

They receive a 5 (and pass quickly) if they have what I consider near-native understanding and speaking abilities. That means I understand every word, grammatical mistakes are very few and not severe, an accent can be present but is not too distracting.

They receive a 1 (and fail right away) if they have a very hard time understanding my simple questions, if they have a very hard time expressing themselves, if they make extremely bad grammar mistakes consistently, or if their accent is so distracting I miss many of their words.

Which factors of the test takers' oral performances made you feel that they were not intelligible?

Language that is so garbled or mangled due to incorrect production of vowels and consonants. Incorrect word stress. No rhythm or linking in their speech. Use of unusual, rare, or innappropriate vocabulary.

When you assess the test takers' oral performances, do you focus on the recognition of word forms and utterances or the construction of meaning?

I focus on both. In phase One, the task is simply to place them into one of three categories, Fail, Continue, and Pass, so there is enough time to step back and make an overall evaluation of the subjects strength as a speaker. For this I include the recognition of words and how they construct meaning and interact with me.

More specifically, what issues prevented you to from understanding what test takers were saying or made test takers' oral performances unintelligible?

I would include all of these examples below could be included. Mispronounced sounds and a lack of Native English rhythm make subjects unintelligible.

For example, some small pronunciation issue causes confusion
Some words mispronounced
Pronunciation difficulty
Pronunciation occasionally unclear
Sometimes pronunciation is not clear, especially at word onsets
others

Could you tell me some specific examples of problematic words, phrases, or sentences the test takers uttered?

No, I can't remember anything specific after so long ☺

In case that the segmental factors caused problems in your understanding, what specific phoneme made understanding difficult? For example, vowel length, or individual consonants?

Vowels often cause a lot of trouble. When subjects make the relaxed vowels tense, it takes a lot of work to reconstruct what the appropriate vowel should have been. V→ W, V – B- P – F causes problems, the palatals are tough, the worst is when people have trouble with the final consonants, as is the case with Thai or Vietnamese students.

In case that the suprasegmental factors caused problems in your understanding, what specific aspects such as lexical stress, intonation, or the relative duration of weak and strong syllables made understanding difficult?

Students who don't put primary phrasal stress in the correct place are hard to understand. They don't indicate contrasts, or ask or answer questions how I would expect.

In the case that your decision of a test takers' oral performance in the phase I is different with the other TA's assessment in the phase II, what would cause this difference?

The subject may have been more or less nervous with one of us, or may have just had a bad performance. Perhaps one of us is a more lenient rater.

Do you think that the oral group discussion format is fitted to measure the construct the EPT oral section is going to assess?

Hmm, I really don't know what you're talking about here.

What do you think the advantages and disadvantages of the oral group discussion format in terms of the construct the EPT oral section intends to measure?

Again, what do you mean by the Oral Group discussion format?

What did you do specifically in the phase I, for example, observing what the test takers discuss, or one-on-one interview?

In the past I would bring everybody to a classroom, and go through the list, calling on people one by one, asking them to say a few words about who they are and where they are from, and what did they think of Champaign so far.

Which format of oral tests do you prefer, and why?

I think Phase II is a good test, though it should be revisitted and modified to adapt to changing trends in the students.
I think Phase I is a good idea, but poorly fleshed out. I think it is unlikely that the raters are applying standardized evaluations of the subjects because there is little training. I also think doing Phase I can be an awkward and uncomfortable situation for the subjects and the tester. I think subjects are not given any information about what the test is, what it is for, and what possible outcomes may be. I think that is a little unethical.

INT: So, What's the purpose of the Phase I of the EPT oral section?

R5: I think…Uhm… this Phase I II… this listen to test taker's English to see if they can pronounce every single word correctly.

INT: So, don't you think… it's going to measure the communicative oral ability or…

R5: The communicative aspect of it…well the question you ask…uhm… let me see…they do mean communicative ability, of course, to answer the questions, because they, these questions are things just we ask them things like uhm…what do you look forward to, you know, about your university life here, uhm, So, they do mean to be able to communicate, but I am not, I guess I am not really looking at their communicative skills. At this point, because, uhm, I guess if you wanna count, like, tone, or you know, intonation, uhm… let's say, maybe things like eye contact, those kind of things as a communicative ability would you count…

INT: Uhm… you focus on more pronunciation,

R5: I would I would I do.

INT: So, where did you get this idea, from your training workshop?

R5: Basically, Uhm, Yeah, in the recalibration, uhm, test supervisor asks us test, tells us to… if there's even one word that uhm we could not catch clearly, uhm, from a test taker, then that person will have to go on to the next level.

INT: Oh, really?

R5: Yes, I think he… yeah, I remember he mentions something like that. So, it's basically listening to each word, and to see if they they can pronounce it correctly? Uhm, because eventually this is a test that is trying to uhm determine whether they need pronunciation courses or not.

INT: The thing is that I am very confused about the rater are going to measure the overall communicative ability, or only pronunciation, or intelligibility like that?

R5: I guess it would be strange to let a student pass, uhm, just because he or she was able to pronounce every single word, uhm, because, uhm, what am I trying to say?, I don't know what I am trying to say. But, uh, it is basically pronunciation that I am looking for at this point. Because that's that I have been told is the aim of this phase, uhm…of oral section.

INT: So, next question, what made you decide a test taker wouldn't exempted from the phass I?

R5: I think this will be a kind of related to the questions number 1. If there were a couple of words, words that were mispronounced or I couldn't catch at all, can't understand at all, then I would have to send them to the next phase.

INT: In that case you assigned the score 1? Or what?

R1: It's score 1 or something. No, no, no, there's no score at this point.

INT: just pass or not.

R1: Yeah, right, yeah.

INT: which factors of test taker's oral performance made you feel that they are not intelligible?

R1: Uhm, ok, they would say a word, well… which I wouldn't be able to understand at all, I would have no idea, uhm, what the word was, so, it's usually, uhm, on the, it's usually word level, it's not, it's not on very well yes, at a sentence level, I

INT: would say. It's just like a couple of words within a sentence sometimes hard to listen to.

INT: When you assess test takers' oral performances, do you focus on more word recognition?

R1:   Uhm, even If I don't understand one particular word that was pronounced by the test taker, from the content of a sentence I can a kind of uhm… predict what that word, I can go back, after I listen to the entire sentence that test taker had just other, I can take the content into a consideration, and then a kind of, uhm, I can I can I know the word was mispronounced. I just, it's just because content helps understand what they are taking about, so there's one one small word that wasn't pronounced correctly, uhm, it is possible to kind of, you know what the word was even even though you weren't able to catch it.

INT: So, then, specifically what issue prevented you to understand what test takers are saying or what factors made test takers unintelligible?

R5:   Basically a way maybe the intonation, the wrong stress in a word, or they may have used a word that was unexpected from like more, uhm, from higher level speaker of English. Sometimes they use they just have wrong word choice, so, uhm, maybe that's one thing that's preventing uhm listener understanding just what was other…

INT: Could you tell me some specific reasons…uhm…I mean can you remember some specific instances that you had trouble with?

R5:   Specific words? I don't I don't remember some specific instances because I have so many students…

INT: So, which specific segmentals or suprasegmentals do you usually have trouble with understanding in your case?

R5:   uhm, lot of times, I notice it's the word stress. The stress of the word, like, for example, uhm, like, let's see, what's longer, "computer" a stress comes with /pu/, like /computer/, and if it's pronounced differently like /computur/, something like that, then it's really hard to kind of understand what that word is, listen to that word. Because I am so used to having a stress on a particular place in the word, that would…yeah that would make it hard sometimes, even though maybe they were using the right word, if the stress was, yeah, in in a incorrect place, then the word make me very hard to understand, and I have to think a while, like, what was that word? Yeah sometimes I figured out, and sometimes I can't.

INT: So, you mean that suprasegmental factors cause more problems than segmental factors? Like intonation or stress?

R5:   Right.

INT: So, I would like to ask if any segmental factors like specific phonemes cause any problem? Or kind of any specific features made you annoying?

R5:   like specific mispronunciation of, uhm…. let's see… I think this Japan on the background of students uhm… like nationality, because I know lots of Korean speakers they have problems with /p/ and… what was it, /f/? I notice this a lot, I know that Chinese speaker sometimes have a problem with /l/ and /n/, that that makes hard to, you know, listen to what they're saying, but it really depends on a person, though. Uhm, /th/ is another, yeah…

INT: Uhm… In case your decision of test-takers' oral performance in the Phase I is different from the other TA's assessment in the Phase II, what would cause this difference?

R5: Uhm, I think it's familiarity, with the like the typical tendency pronunciation tendency, it's like people from different background have, for example, I was living in Dubai for two and a half years, so I am really used to listening to Indian English. So, I have no problem whatsoever listening to that. That pronunciation, however, sticks to their accent maybe. I'am just so used to it. Sp, I wouldn't have no problem with listening to that type of English. But, other people who are not used to it, who never been used to this kind of English, I think, may have more problems, listening, and tryi… catching every single word.

INT: How about Korean speakers or Chinese speakers? Do you think you are familiar with students with these nationalities?

R5: Uhm… not necessarily. Because until I did the tutoring for uhm… for 89, you know, we do pronunciation tutoring, I didn't know that Chinese speakers have a problem with /l/ and /n/. I would never have guessed, but those two sounds are difficult for these people. So, it really is, yeah, a matter of experience. But, I don't necessarily, I don't think, I don't see myself necessarily, having more ears with Koreans or Chinese speakers.

INT: Then, go on to the format. We usually used the group format in the Phase I, right? Do you think that the oral group discussion format is fitted to measure the construct the EPT oral section is going to assess?

R5: Uhm, if we are going to use the oral group discussion format, then I think all raters have to use it. It can't be one rater is using the individual interview style, and then aonther rater in the next room is using a oral discussion. Because it's totally if students have to speak in the front of a group, it totally changes dynamics. It all make a lot of students nervous, and it will be hard, because everyone's in the room. You, you are not able to, raters are not able to focus on one student's pronunciation. We can't be as focused on one student as we like to, so, I really, I haven't done actually the group format. So, I can't say how it might be different from, actually, the individual, uhm, format. But, I think they're two completely different things. So, we are, we decide to do the group discussion format, I think that everybody should do that in a same way. Or, yeah.

INT: So, you prefer individual interview format?

R5: Yeah, I would prefer one-on-one one. Because listening very closely to the pronunciation, so I think it's important that you know there's no other distractions, not not a lot of people. If there are a lot of people present in the room, it might be hard to understand what is a student trying to say, so, definitely I I would say the, yeah, just one person in the room.

INT: This might be a very similar question, though. But, can you come up with any advantages or disadvantages if we are administering the group format in particular in terms of measuring intelligibility?

R5: Right, right. And then, we can't… It's it's really hard to make sure that every single student have has an equal amount of time takes to respond they want to, yeah, I would say, it could be unfair, unless raters are trained properly for it,

because we currently aren't trained to do during the recalibration there's no mention about the oral group discussion format, so…

INT: Thank you. We 're done for now.

What's the purpose of phase I of the EPT oral section?
It is to screen out students who will receive either a level 5 or 1 from the oral interview based
on their 3-minute speech sample.

What made you decide that the test taker wouldn't be exempted from the oral section phase II?
It depends on their pronunciation, whether they can accurately pronounce words and have good intonation and stress.

Which factors of the test takers' oral performances made you feel that they were not intelligible?
If the test taker's pronunciation was difficult to understand and hindered interpretation. I also
look for accents.

When you assess the test takers' oral performances, do you focus on the recognition of word
forms and utterances or the construction of meaning?
Because the purpose of the oral interview focuses on pronunciation, I focus more on the recognition of word forms and utterances. It all depends on the intelligibility of their utterances depending on the pronunciation ability.

More specifically, what issues prevented you to understand what test takers were saying or made test takers' oral performances unintelligible?
 For example, some small pronunciation issue causes confusion
Some words mispronounced
Pronunciation difficulty
Pronunciation occasionally unclear
Sometimes pronunciation is not clear, especially at word onsets
others

--pronunciation is unclear, intonation is bad, linking, stress, mispronunciation

Could you tell me some specific examples of problematic words, phrases, or sentences the test
takers uttered?
I don't recall specific examples from Phase I, but I remember that students had trouble pronouncing words like 'vinegar' 'soporific' from the readings in Phase II

In case that the segmental factors caused problems in your understanding, what specific phoneme made understanding difficult? For example, vowel length, or individual consonants?

Vowel length

In case that the suprasegmental factors caused problems in your understanding, what specific
aspects such as lexical stress, intonation, or the relative duration of weak and strong syllables
made understanding difficult?

Intonation

In case that your decision of test takers' oral performance in the phase I is different with the
other TA's assessment in the phase II, what would cause this difference?

Experience, awareness of the purpose of the oral pronunciation class.

Do you think that the oral group discussion format is fitted to measure the construct the EPT
oral section is going to assess?

Yes, because the EPT oral section does not test fluency, it tests pronunciation. Oral discussion
can provide a chance for students to participate in a class-like environment where they can introduce themselves and share ideas on general topics. The underlying construct of the writing section of the EPT is to simulate an ESL-like classroom and I think the group discussion
simulates that.

What do you think the advantages and disadvantages of the oral group discussion format in terms of the construct the EPT oral section intends to measure?

Advantage: Saves time, lessens test anxiety, simulates a classroom setting
Disadvantage: students might be familiar with doing introduction, so may produce better pronunciation compared to spontaneous questions on general topics

What did you do specifically in the phase I, for example, observing what the test takers discuss,
or one-on-one interview?

I led a group discussion and asked students to introduce themselves. Then I asked them general questions outlined in the interview guideline. Students were in a group, but individually administered.

Which format of oral tests do you prefer, and why?
It depends on the number of students we have. On large days, I definitely prefer doing phase
1 and 2 separately for practical reasons. On smaller days, I like doing Phase I and 2 at the same time because I think it might seem like a more natural way of interviewing, a format that students might be more familiar with.

Definition of intelligibility:
-- The definition of intelligibility is difficult to pinpoint because I think that intelligibility is a combination of word or sentence recognition by the listener and whether the pronunciation of the speaker hinders understanding.
In intelligibility, I don't think you can say that the listener or the speaker is more important because the interaction between the interlocutors is essential in negotiation of meaning. The speaker needs to be clear in word choice pronunciation. The listener needs to be in tune with the speaker. Because a conversation is an interaction, both interlocutors are equally important.

Where did you get this idea? Trained? From class?
My thoughts on intelligibility came from two sources, EIL 488 and the experience from EPT rating. The EPT interview rating was the first opportunity that I had in assessing test-takers' pronunciation and it was interesting to see how classroom knowledge transferred to testing context.

Do you think we need tests to measure overall oral ability or any ESL speaking course to enhance a whole rage of oral ability?
By oral ability, I am assuming that you are referring to communicative ability. I feel that EPT test-takers as well as many international students want a class that focuses on oral ability. However, I think a more helpful class might be where students are taught skills in academic/professional speaking, such as leading discussions, presenting papers, or proposing research. I think a course like this would be a better fit in a university.

When you assess test takers's speech in the EPT oral section, their grammaticality affects your rating?
No. The focus of the ESL pronunciation class is pronunciation so the focus of assessment in ESL oral is pronunciation. Grammatical accuracy or fluency is not important in getting a good score in EPT oral.

Group discussion format
Advantage: Time saving (!!!!!!!!), lessens test anxiety
Disadvantage: There is the fear of copying another test-taker's response, but because we are only looking for pronunciation ability, that shouldn't be an issue. Students might be embarrassed to talk about themselves in a group environment.

I prefer the group-administered oral because it saves a lot (!!!!) of time. The primary purpose of the phase I is to save time and I feel that the group oral format does what it is supposed to do. In addition, even if students receive hints from previous students, because the purpose of the assessment is pronunciation, I don't think it will affect the results. If the raters are properly trained, it should not be a major issue.
I think it might be interesting to combine oral and writing assessment into the same session. Because the writing test has a discussion session, if we can find a way to rate students' pronunciation during that time, the integrated format might stimulate a more lively discussion.

R7:   Sorry, I just remembered something I had in my hands.

INT: So, what do you think the purpose of Phase I in the EPT oral section?

R7:   Phase I is just the, just the three minute interview.

INT: Yeah, oh yeah. The, in the morning, you have the interviews, so what's the purpose of it?

R7:   The purpose of it, the purpose of the interview is to sort of determine if there are people that you can eliminate right off the bat. After talking for about 3 minutes, their speech is completely intelligible. There might be, they might stumble with words or grammatical things, but you're just looking at the pronunciation, can I understand every word I'm saying, or are there things that are unclear?

INT: Uh huh. So like, it's not about, from my understanding, it's not about overall oral ability. It's about . . . .

R7:   No.

INT: It's about, right, pronunciation, or intelligibility.

R7:   Yes. Pronunciation. And the students know that. Before you, everyone's out in the hallway, kind of, and I would say to everyone, don't worry about your grammar. Don't worry about how you're organizing it. I'm just paying attention to your pronunciation, and just listening for how you speak, the way that you speak. So they're, they know that for the portion of the interview, or for that portion of the day, the focus is pronunciation.

INT: Oh, so you explain what it is about.

R7:   Mmm hmm.

INT: Because somebody said that it's kind of a secret.

R7:   No. I tell them, we tell them what we're listening for. We don't tell them, if you come back in the afternoon, that means you did a bad job. But most of them figure that out, because by the third day or the fourth day, people who would come in the afternoon, testing Phase II, would say, oh, if I had interviewed them, they would say, oh is it because, "what did I do wrong?" And I would, we usually just say, we just didn't get a large enough speech sample, or we just would really like to hear a little bit more from you. It's just sort of a nicer way of saying, you know, that maybe there were things we didn't understand.

INT: Yeah. So it's about kind of, so you know, pronunciation or intelligibility. So how do you define intelligibility?

R7:   Well, for me it was, what Professor Dickerson had told us was, if there are any words that you don't understand, have them come back for the afternoon session. So the first day, I had a ton of my students come back. I would say maybe two thirds come back for the afternoon session. But then when I gave them the afternoon test, they really did well enough on it that they didn't need the transition course. And you can make a certain number of errors, maybe like less than 30 errors or something, and you're not even considered for the course. So I kind of realized that I had higher standards than the other raters for Phase I. So I sort of adjusted mine to be a little bit more lenient in the Phase I interview. I think the first day I was, and also some of, a lot of the other raters were, spoke East Asian languages, so they spoke Japanese, and they spoke Chinese. I'm a lot less familiar with those languages, so for me maybe, maybe when you would hear someone

making a mistake but you speak the same native language, you know what they're trying to say right away. You don't even realize it.

INT: Yeah.

R7: But for me, it was, it seems more like a mistake, because I didn't know what they said.

INT: Okay. Actually, I think that you only know that because of an operational definition about intelligibility. So there's no agreement about that, so. I would like to know which definition is most closest to your understanding: one word in an utterance and recognition and native speakers' perception, or adjustments in a rating scale, how difficult or easy is an utterance to understand ?

R7: Like I was saying, I think it is kind of colored by, or affected by, my own perception as a native speaker and a speaker of English and Spanish and not a speaker of East Asian languages. But I think intelligibility would sort of be, I kind of agree with the third one the most, I guess. It's like, it's sort of a rating scale. So I would say it's a continuum. You know, not yes or no, but how intelligible is this person? How difficult or easy is it to understand what they're saying? So it's sort of like a scale.

INT: So intelligible like accent, or grammar things. So do these factors have kind of an effect on your rating or your intelligibility?

R7: Right. That's what I'm saying. I think that like, for example, some that I tested in the afternoons, Phase II, that means they didn't pass Phase I or they were questioned, was an Indian speaker. And I thought his English was great. But I think maybe he was interviewed by someone who was less familiar with Indian accents. He made very clear as an exam. It was kind of like, he was one of the best people I saw going through Phase II. And so I wondered, was he just matched up with someone who didn't, wasn't as used to their accent, you know? Like the way I felt with the Asian accents on the first day, the East Asians.

INT: Yeah. So how about grammar? So if somebody's grammar is very bad, and is it, I mean, that kind of stuff a factor in your understanding or their intelligibility or?

R7: Does bad grammar hurt the intelligibility?

INT: Yeah, yeah. Or the score in the Phase I?

R7: Um, no, not at all. The grammar doesn't affect it. But I do think in the real world, as far as intelligibility, that people who haven't studied ESL and aren't sympathetic to that maybe, do judge grammar maybe more harshly. Grammar and accent, I would say, are the two things that like, this person clearly hasn't been in this country a long time. Or they start to stereotype. But with Phase I, no, you're only looking for can you understand the words they're saying.

INT: Yeah, so not grammar. So it's kind of a similar question, but when do you assign a score of 5 to the test taker, and then when do you assign a score of 1 to the test taker?

R7: I mean, a lot of times you could tell within the first minute, but you would let it go for another minute or two just to be sure. Basically, if there was one word that I thought, oh, I'm not sure if I heard that quite right, I would sort of let it go for another minute or two, and usually either I was like, oh, no, this person is pretty good. Or they would make a few more. Like, you know, say a few more words of things that I completely didn't understand, and that's when I would give it the 1.

INT: Yeah.

R7: So, like for example, I remember, I don't know why I remember this. One kid said, "scolersheep" [note – not sure how to spell the way she pronounced it]. And he

was saying "scholarship." And at first I didn't understand it initially. After I thought about it for a couple of seconds, I understood. But we were supposed to process it as, if we can't immediately understand the utterance, then it's unintelligible for the purposes of Phase I. So, but then in the next two or three minutes, he didn't say anything else that was unintelligible. Everything else was perfect, so I thought, okay, it's just a problem with the "cha sha" kind of sound. It really doesn't need to take a pronunciation course, just one little consonant blends.

INT: Yeah. So, yeah. I think so. So I just like to know, what factors, what factors are apparent to you from understanding what the students are talking. Like how about intelligible suprasegmental factors or segmentary factors. What factors are more important or more harmful for your understanding?

R7:   I'm sorry. What kind of factors?

INT: Yeah. What kind of factors are, suprasegmentals or segmentals, like the overall intonation or . . . .

R7:   Are you talking about like . . . .

JEL   Vowel, consonant, like that.

R7:   I mean, intonation is something that you do test on Phase II. It is on the exam. But I only looked at it as, is the intonation so confusing that I can't tell what they were trying to say. So like, which rarely happened. It was really more, I think, certain consonants and things that, or vowels, that sort of distorted the word. Or stress, word stress, that to me distorted the word beyond recognition. So, and then I would consider it unintelligible. Yeah. I can't really think of . . . .

INT: So can you rank it? Like somebody said that like word stress is most problematic for their understanding, but that kind of thing.

R7:   Um, I think it depends a lot on the speaker. Sometimes there were clear problems with word stress. With other people there were just a couple phonemes, but a couple commonly used phonemes, so they were talking, and it was just, after a while I began to see a pattern. Oh, this person's really struggling with, you know, "rl" or with theta. But I think that the word stress and vowel consonant problems are the easiest to hear agreements.

INT: Yeah.

R7:   If complex intonation things, like they might not have asked you a question. They might not have needed to use specific intonation patterns the way that in the Phase II interview, like they have to ask a tag question. They have to, we do like a little conversation, like a role play. And so you have to use specific terms and patterns, and then you'll see that there's certain ones that almost all students get wrong. So regardless of really their pronunciation ability in other areas. So like they could have been great with all the vowels and consonants pronounced correctly and then still the intonation was off.

INT: Yeah, yeah. So it's like, it depends on the test takers, yeah. Okay. When we, when we limit the content the programs only to like segmentary factors. I mean, consonants or vowels, which one is more problematic to you? So vowels bother you more than consonants or some special consonants make you understand, hard to understand?

R7:   Um, again, I think it's just sort of what you're more familiar with. So like when I would be working with a Spanish speaker, I would immediately know what they were trying to say, just because I know. Like in Spanish, you can't start a word with a consonant blend like /sp/ So like they would put, they would add a schwa vowel at the beginning or something like that. So like, but I'm very attuned to,

that's what they're going to do. Like this is a very common problem. So like I said, I think it depends a lot on the native language of the speaker, whether they're going to have maybe more consonant, or more problems with consonants or vowels.

INT: Yeah.

R7: I don't know. I didn't specifically notice that, that people had more problems with vowels or more with consonants.

INT: So they kind of, they kind of to specific examples, specific common problems. Do they have an effect on your scores?

R7: On the scores I gave them?

INT: Yeah, yeah.

R7: Mmm, I don't think so. And I actually, most days I had last names W through Z, so I had almost all East Asians. Almost all China, actually. Because a lot, like Spanish last names don't really end in "w," or "z" they can, but it's very, it's rare. So because of the group I was working with, I think it wasn't maybe representative of the larger group. So it's also that.

INT: Okay. So, can you remember specific example that any kind of words or sentences you couldn't understand?

R7: Specifically?

INT: Yeah. It's kind of hard, I know.

R7: Um, no. Sometimes I would think it was a word that I didn't understand, but it was really a place name or something in their country. Because a lot of the questions were like, where is your favorite place to vacation in your country? Where would you want to live in your country? What are the most important national holidays in your country? So a lot of times they were talking about a tradition or something, and I would assume it was a Chinese word. It could have been an English word I just didn't understand. So I think maybe, those are nice topics to work with, because it's easy for them to talk about, you know. Like something that they're very familiar with. But maybe it's harder for, to understand, or to know for sure from what language they're talking in all the time.

INT: Okay. So you already said about kind of misplacement. I mean like in the Phase I, you gave some students to score 1, but they can be, they could be exempt from Phase II. They might like, might have take some like three or four, so what might be the cause for the kind of misplacement, a different rating score?

R7: I think, for me, I thought I was following the directions. If there was one thing, I did not understand, I sent them to Phase II. But the thing is, if there was only one thing I didn't understand, they probably were going to pass Phase II without any problem. They probably weren't going to need a pronunciation course. Where, and so basically what I did, instead of sort of following the exact directions, I just aligned myself with what the other raters were doing. So kind of the way other people were looking at it, I guess. And then, and I think it worked out better. We still, still a lot of the people that came through Phase II ended up, I think, not needing the pronunciation course. The other issue, like Professor Dickerson tells in advance, we've been having higher and higher numbers every year not need a pronunciation course, because they're coming in with a higher level of English. And it's because he said initially, like maybe half only will need a pronunciation course. I think way less than half end up taking the pronunciation course. So that sort of, I kind of had that statistic in my mind, thinking well about half are going to need this course. And then when I would test them for Phase II, I would realize, you know, no. Not quite half.

INT: So I see. So that's kind of redundant question like . . . .

R7:   suprasegmentals

INT: Yes. suprasegmentals so the, about lexical stress, intonation, or relative duration of weak and strong syllables, like. So which one is the most problematic for your understanding?

R7:   Um, again, when it's the wrong intonation, I generally still understand. So I would say, word stress, which is lexical stress, which is also related to weak and strong syllables. Their pairs intertwine, because a lot of the times they'll make the wrong stress on the word because they'll lengthen what should be a reduced syllable, something like that. So again, when, it's probably a combination of those when they either make a strong syllable weaker. More often they make a weak syllable strong. And then there's, additionally, the problem of stressing the wrong syllable. Which, and that was a very hard section for students on Phase II, the part where they have to read sentences with unfamiliar words. So they're really difficult words. But native speakers still sort of know how to pronounce them even though they've never heard them. And that was the section that a lot of people struggled with. So I think that is sort of a problem area.

INT: Yeah. Okay. I see. So you have more concept of like intelligibility or understandability as from kind of a specific class, like for phonology class or some training station about EPT?

R7:   Right. So I was looking at intelligibility as far as the instructions we were given about how to determine whether or not these students, first of all, need to come back in the afternoon and take this additional exam. And just secondly, like I don't know. Basically I feel like intelligibility, as far as what a normal, non-ESL student would think is a little bit different than for our purposes. How, I guess we were being more stringent on intelligibility, because if there was even one or two things that you couldn't fully understand, you wanted to make sure the person didn't have other, deeper problems of phonemes and things like that.

INT: Yeah.

R7:   And I even gave advice. Like sometimes on Phase II, a student would do really well, but they would have a problem with one specific thing. And I would say, this is just, this is your area that you need to work on. I think they appreciated that, too. But it's not really, like they didn't need the full pronunciation course. There was just a couple specific problems.

INT: Yeah, that's true.

R7:   So is that person still intelligible? For the most part. And again, they go to choose, or not choose, but they were given a topic. But they can say whatever they want to say. So some students could have just avoided problems that they know they have, words they don't know. Obviously they're not going to use words they don't know when they're talking. So I think there's, like in Spanish, I'll just say things that I'm comfortable saying, that I've probably already said to someone before. So you're pronunciation may seem better than it actually is. It's not a factor.

INT: Yeah. So I think that, I'm officially blended, but acquiring some segmentary factors like some problematic vowels or consonants are easier to acquire, like segmental factors. So the, speaking English we give proper intonation or proper lexical stress. It's very hard for me. So I think that it's really hard to learn, that kind of suprasegmentals. So do you, were you informed about some specific format you had to take in the Phase I? I mean the . . . .

R7:   We were given the sheet with possible topics for discussion. And so I never really used the same one. I usually used about four or five that I felt like worked really well. Because there was one that was like, describe your family. And I didn't use that one, just because, like I said, most students have probably described their family in English a hundred times. So I think they're going to say to you, they probably will pronounce it better than their true pronunciation, so I kind of wanted them to have to talk about something that they don't always talk about. One question I asked that worked really well, and I can't remember if it was on the sheet or if I just asked it, was, a lot of them had only been in Champaign for a few days. So I would say, what have you done since you've been in Champaign? So they would have to say, well, okay, I got here, and first I went to, you know, a hotel because my apartment wasn't ready. And then, so they would have to like list everything they had done in 3 days. But other people, like I said, I said describe important holidays or other aspects of your home country. One of them was about superstitions in your country. So I had lots of Chinese students tell me about the color red and certain numbers and things.

INT: Oh, yeah.

R7:   But like sometimes a student wouldn't be able to talk for the full 3 minutes, so I would ask a second question. But usually they could talk the whole time.

INT: So you could, like one by one or?

R7:   Right. So what I would so is, they're all out in the hallway, and I explain the directions. I say that we're going to do with approximately a 3 minute interview. I am focusing on your pronunciation, the way that you speak. I don't care about your grammar. I don't care about anything, those other things right now, just pronunciation. I'm going to give you a topic, and I want you to try and talk for the full 3 minutes. So I'm going to try and just listen. If I need to ask another question I will, but hopefully I will just be listening. So that they know that they should be continuing the conversation on their own. It's not like an interview. It's just like I ask you a question and you sort of present.

INT: Oh, yes. So other peoples listening to what the other test takers are saying.

R7:   Um, they did that in some classes. I didn't do that. Just because I thought it was kind of distracting and things. So I had them say it alone. Plus, some people were in a lecture. I was in a small classroom. So, um, I would call the one student. And when we finished, I would say, okay, give us the next name. Can you go out in the hallway and get this person? So I just stayed in the classroom and the students would file in.

INT: Oh. Okay. So because the, these days, the, the EPT faculty advisor said that we kind of trying to employ the group format. So, yeah.

R7:   Right.

INT: So they, the test takers are making a group in their one classroom, and then they are kind of listening to the others. Yeah.

R7:   Right. I was afraid that that might make the student who's being, who's talking nervous. So that can maybe affect your pronunciation if you're, not your pronunciation, but just like the fluidity of your speech, if you're very nervous. And also, I thought like, because the prompt should be not something you're expecting. You know, if I wanted to give them all the prompts in the hallway, I would have. But I wanted them to talk naturally. So if you've just heard someone, there were only like 10 questions on the worksheet. And I only ask usually four or five of them. And so if you already heard all those students give a response, you can start

preparing your own responses. Then when you're preparing it, then I feel like you have more time to think about, oh, okay. How would I say this? Oh, I don't know how to say that word. How can I get around it? [laughs] And so I didn't want them to have that time to think about, okay, like circumlocution and other strategies to get around it. I wanted it to be sort of a natural, on the spot response. So that's why I kept them in the hallway.

INT: Uh huh, yeah. Oh. So, but because we have a very limited time, and limited money, and the whole population who has to take EPT is sharply increasing in this university. So we have very small raters and lots of test takers. So the university is planning to employ the group discussion format, like group format very actively these days. I heard that. So what could be the advantage or disadvantage of the group format. So if they are sitting in a group, and you can interact with one student and another like that, and other people are seeing what you and the other test takers are saying. Or they can interact each other. And . . . .

R7: Um, I can see, I can see an advantage to have two students interview each other. So I would sit there and listen to two students talk. I think any more than that, this is another reason I kept them in the hallway. Is because the students start to talk among themselves, and then it's distracting, because you're really trying to focus on every syllable they're pronouncing. And so I didn't want that kind of noise. But I also didn't want them sitting there listening. So I kept them in the hallway. But if, to have two people at once I think would be very, would be effective. What I would say is I'd give them a topic, and the two of them would have to talk. If they have very different proficiency levels, it might be problematic because, you know, one student would maybe prompt the other one, and the other one wouldn't give much in response. So where I feel like maybe I, I might be better at prompting, if the student runs out of things to say, I would kind of quickly come up with, okay, "well let's talk about this." So that might be hard to do if I'm just sitting back, sort of. But yeah, I think an interview could work. It's just that you would have to be judging two people at once, which is a little bit more complicated for the rater. And I think there would be a tendency to sort of rate one against the other. [laughs] So you would think, well, this one was better than that one, so I'm going to give the worst one a 1. And maybe that wouldn't happen, but it's, I think, a possibility. So you have to look into controlling.

INT: Okay. I see. For, so, I was wondering that we assess the only pronunciation rather than the whole communicative ability. Like the whole range of oral abilities. So do you think we need other test to assess or measure the communicative ability or like overall oral ability for academic courses in EPT?

R7: I think if you were to do that, you would have to change the pronunciation course, 110, 510. Because I think right now the Phase II exam reflects the content in that course. So if you struggle with this content, you will benefit from this course. It's a direct relationship. So I think if you're going to change the way you test, and you're going to ask for more, you know, communicative competence, the course should also reflect that. Um, I don't know. I think the reason pronunciation is so important on this campus is a lot of these people will be TAs. And if you're going to be teaching a course and your students can't understand you, that's, you know, a big problem. And also in your field, you know. You want to be understood. I mean . . . .

INT: Yeah. So if the pronunciation is one of the biggest problems to communicate with international students?

R7:   Well, because of what, if you were talking about communicative competence, what more are you really going to include? Like how learners use other strategies? Like gesture. I mean, those are helpful when you're, for you to use again, like in every day real life, if you can explain something using what you've got. But I think at this level of academia, if you're in a master's program, a Ph.D. program . . . .

INT: So we don't need any, any, like the very good, any program like Ph.D program doesn't require any, much communicative competencies. You mean like . . . .

R7:   I'm just, what, okay. So what do you mean lack of communicative competency. What else would be the tests?

INT: Like, how about the communication classes, what they're so sure of, what their opinion, like make a presentation. Or like communicating with the other, the other students or other professors like that.

R7:   I think that would be very difficult to measure and control. So, I mean, even pronunciation can be difficult to think, oh, did they say the right intonation that time? Like, because they read everything twice. You only grade the second time. Uh, I just, I mean, yes. It's good if you have a high level of communicative conscience, but I'm not sure how you would directly assess that.

INT: Like TOEFL IBT speaking session.

R7:   What is it, IBT?

INT: IBT speaking session.

R7:   I don't know what the IBT is. [laughs]

INT: I know. So, like SPEAK test. No idea. Okay. So yeah. I understand what you are saying. We're almost done. Okay. This may be all. Thank you.

INT: Alright. Yeah, it's working. [noise] Sorry.

R8:   [laughs]

INT: First of all, you are a native speaker of English?

J   Yes.

INT: So what's your major right now?

R8:   MATESL.

INT: So what, what was your major when you were undergrad student?

R8:   Advertising.

INT: So, and then how many years or how long have you taught English to secondary speakers?

R8:   This is my second year, so this is my third semester teaching.

INT: Mmm hmm. So do you have any other experience to teach other language rather than English?

R8:   Um, no. Not languages. I was a substitute teacher, before getting into grad school for elementary school children. But that wasn't any particular subject. It was general.

INT: Yeah, general.

R8:   General school stuff.

INT: So have you taught any English pronunciation course or?

R8:   I've taken one. I've not taught it. And part of the course that we took was tutoring in pronunciation. So that's like . . . .

INT: Ah, yeah, yeah. For ?

R8:   Yeah, yeah. That was the extent of pronunciation teaching on that.

INT: Yeah. So this is the first time when you rated the speakers in EPT section?

R8:   Right.

INT: Yeah. Okay, that's good. What do you think is the purpose of the Phase I?

R8:   Phase I?

INT: Yeah. Of EPT.

R8:   That is to, how should I put it? Well, judge initially their speaking ability and to kind of weed out the people that obviously don't need pronunciation class. And to really identify the people that, who need more testing to see if they do need the pronunciation class.

INT: So, so how do you define intelligibility?

R8:   Um, well, Dr. Dickerson told us to, if we could understand every word they said and not have to stop and think, what was that word again. Oh yeah, this word. Then that means that we could understand them well. And I think for the majority of people I interviewed, I really had no major trouble understanding them. I don't know if that would be so with some of the other raters, or like general people on the street. So I don't know if it's because I've worked a lot with international students, and my husband's international.

INT: Oh, really.

R8:   So I'm used to hearing like the different accents. So I don't know if that influenced a lot, but, especially towards the end of the week in the Phase I, it got a lot easier to weed out people that I knew would not need pronunciation.

INT: Yeah.

R8: Even if they had one word that I had to say, "what was that word again?" The 2,000 other words they said were pretty good. So, yeah. [laughs]

INT: So I'm kind of a little bit confused about that, because the test supervisor said that the understanding the every word. So I mean, like just think about this kind of case that you understand the every word. But you cannot get the oral sense, I mean. So like in case that you didn't make any sense what they are saying, but you can understand every word they uttered. In that case, can you _____ (4:30)?

R8: Yeah. Like if they use the wrong words but they say it correctly kind of thing, but don't know what they're talking about things?

INT: Yeah.

R8: Um, yeah. I'm strictly listening for pronunciation, really, at that point. So even if they're talking about their major and I have no idea what they're talking about because it's crop science or something, as long as I can understand the words they're saying and they're not throwing in a word like crop science, crop science, cat. [laughs]

INT: [laughs]

R8: And it kind of makes sense. Then I would pass them.

INT: Oh, yeah.

R8: Yeah. I didn't run into any one that was so beyond my comprehension that I didn't know what the heck they were talking about. Luckily.

INT: Oh yeah. So like, any heavy accent or any very bad grammar, like can just you understanding that?

R8: Um, yeah. There were a couple of people who very, very heavy accent, or people who were trying very hard to pronounce things well, but would hesitate so often or maybe stutter a lot trying to get it out, that it would really impede my understanding of them. So those people I would recommend for further testing, Phase II.

INT: Ah, okay. I see. So the concept of intelligibility, is it just limited to the rating in the EPT section, or it apply to your oral understanding, I mean, oral intelligibility in the real world?

R8: Um, I, for me it was just the EPT. Because I know, and you know, fellow teachers who haven't dealt with a lot of international students have said, I cannot understand them. I just have a hard, I can't pick up that accent. Yeah, there's a lot of people in the real world who would probably be like, "I can't understand you at all." But with the experience I've had, and knowing that I'm listening for school type EPT only, probably a lot more people pass through than possibly should have. Everyone could, not everyone, but a lot of people probably could have benefited from the pronunciation classes. But we don't have that many sections of those. So I mean, you can't have like 500 people in a pronunciation class. So it naturally shakes out...

INT: So, sometimes when you assign the scores, you gave them like pass or continue or fail, right?

R8: Yes.

INT: Yeah.

R8: Yeah, yeah. That's what, it was 5, I think, was a pass. It was a rating of like 0 to, 0 and 5, I think it was. Now I don't remember. Oh gosh. [laughs]

INT: Yeah. [laughs]

R8: Yeah. I think that like the zeros are plus and minus or something, and minus for the people that needed to come back for the Phase II. And the pluses were the people that passed through, didn't need it.

INT: Just passed.

R8: Yeah.

INT: Yeah. So when you understood every word they said, then you give the pass score?

R8: Yeah.

INT: Then you didn't understand every word they said, then . . . .

R8: Then I would have them in the afternoon, right.

INT: Yeah. Okay. So, I want to ask about some accent more. So do you have any familiar intonation accent?

R8: What?

INT: I mean, kind of, in my case, in my case, I'm very familiar with Korea speaker of English. So even if the Korean speakers has a very bad, bad or very heavily accented, but I can understand them well. So this kind of speakers, I mean.

R8: Yeah, yeah, I do. With the Chinese speakers, because my husband's from Taiwan. And a lot of my students are from China, too, so I'm familiar with that. And Japanese speakers, because I've studied Japanese. I usually have a better time understanding them than say, people from Spanish speaking countries. Sometimes I have difficulty with that. And usually I think, like Turkey and the Middle East, I tended to have more problems understanding them, too.

INT: Oh yes. So that kind of, that kind of a familiarity with some specific accents has kind of effect on your score or rating?

R8: Yeah, I think it did.

INT: Yeah.

R8: I think it did. Because I had, I don't know all the time if I really understood all the words clearly, but in my head, because I'm familiar with the accent, it seemed like I did. But if I looked at tapes later, I'm not sure I could actually say, "yes that was each specific word." I hope so, but [laughs].

INT: [laughs] Okay. Yes. So I just like to know the some specific factors like, you know, like super segmentary factors, like you know, intonation or lexical stress like that. Whereas segmentary factors like consonant or vowels. Which one makes like, which one most contributes to your intelligibility while you are rating the test takers on intelligibility?

R8: Um, let me think. I think a lot of it, the vowels had a lot to do with it. Intonation, for the people that I interviewed, I didn't have intonation as much of a problem. I'm not sure about other backgrounds, if the intonation would be different. But the people that I interviewed seemed to use stress and rhythm and that pretty well, in my estimation. Um, the vowels, some vowels and some like blends, consonant blends at the ends of words would be hard sometimes. Like the "th" sound would come out "s" or something. Or the vowel contrasts, like "e" versus "uh" contrast. Those would affect intelligibility for me.

INT: Yeah. So you mean like segmentary factors are more problematic to you than like super segmentary factors like word stress or intonation things?

R8: I think so. Because I don't remember many of the stress intonation problems. They might have had them, but I don't remember them as clearly as I do like the vowels and the consonant blends. Yeah.

INT: Okay. I see. So the next question is about some kind of a misplacement. I mean, like in the Phase I, are you, when you gave score 1 or fail, so they had to come back to Phase II in the afternoon. But they find they could be exempt from Phase II. So they didn't need to take any pronunciation courses. So it's kind of the

difference, difference about the rating, and from the different raters. So what could be or might be the cause?

R8:   I think because Phase I, we were, it was stricter. As in if you couldn't understand every word they said, you moved them to Phase II. So it got more strict, even if 90% of they said, you could understand. If you still couldn't understand the 10%, you moved them to the afternoon. And the afternoon, the tests are much more rigorous in that they need a certain score. They need to be below that to get pronunciation. And most people, once they got more comfortable talking and had spent the whole day involved in English, I think they kind of improved.

INT: Oh really? [laughs]

R8:   Yeah, like I've been hearing it all day. So now I'm more comfortable with it and speaking better in it. I don't know if that had an affect on it or not. But, yeah. I was not surprised that a lot of people that we sent to Phase II would pass Phase II and not have to take the pronunciation course. I think they were, not very few, but there were specific cases, I think, that really needed pronunciation. And others that were kind of like, mmmm.

INT: Yeah, just, yeah.

R8:   Borderline or even. Yeah. People who didn't get the required number of marks on the test to take pronunciation, but in speaking with them, we as raters were like, you know what, pronunciation would really be good for these people. And we give them recommendations for that, so.

INT: So, yeah. Yeah, I think so.

R8:   [laughs]

INT: So how did process the Phase I? I mean . . . .

R8:   How did it work?

INT: Yeah. So, I mean, you call the test takers one by one or?

R8:   Mmm hmm.

INT: I, I know that like you are encouraged to make some groups and . . . .

R8:   For, not for Phase I. We weren't, no.

INT: Yeah, so . . . .

R8:   Yeah. Phase I was, well, let's see. We would have a group of names, and we'd take them all to like a classroom, a separate area. And then we'd talk to them, usually one by one in there. I guess, I'm not sure how the other ones did it. But I would sit, usually, away from the other students and call people over to me to talk. And it would be, "hi, how are you? What's your name, where are you from?" And then I had a list of different questions to ask them that I would select randomly from. And I tried not to have too many people, like too repetitive, so the third one would be like, oh she's going to ask me that third question. I'd better prepare now. [laughs] I would try to vary it a lot more.

INT: Yeah.

R8:   So, um, I don't know if other people had like groups where they could have a little group discussion. I suppose that would work too, but for me I just called them up one by one.

INT: Oh, really. Because I heard that there are so many people these days, so the, just the rater makes a group, and then you can process the interview one by one. Then the other students are listening to what you and the other student are talking. So you didn't take that format.

R8:   No, no, I didn't really. It was, the other students could still hear a lot of our conversation. And I think once or twice I had people outside the classroom and

then coming one by one. But yeah, mostly it was just, it was just strictly one-on-one conversation for me.

INT: But actually the… we call it like a group administered format or a group discussion format, so it's very newly innovative format in this speaking assessment. So this university so encourages the raters take that kind of format. But I just want to know, if you can take that kind of format next time, what would be a disadvantage or an advantage? Can you like it or?

R8: Yeah. I think, it does have some advantages, I think. I think the students might feel a little bit more comfortable knowing they're talking with other students too, not just me, who they know I'm evaluating them. And also for me as far as comparison, if I hear two or three speakers responding kind of at once, then I can kind of see, "oh, he speaks a lot more clearly than he does." Or, "wow, she really needs to go to Phase II, because I'm not understanding her like I am him." But for disadvantages, if the group is talking and they start talking and talking over each other, I have a lot less control of what I can hear and what I can pick up. And even time limits. And if I ask a question, and they say one statement each, and then they just sit there and think like, okay, we're done now. It might take longer. I don't know. But I think that there are advantages to that too.

INT: Yeah, so you think that you can rate them, and then you can get enough speech example from group discussion like that?

R8: I think it's possible.

INT: Yeah, yeah, it's possible. Okay, so . . . .

R8: Small groups. [laughs]

INT: Yeah. The size matters, I think, yeah. Then actually, in the EPT speaking section, we focus on only on pronunciation. A little bit wider sense, that's kind of focus on intelligibility. So do you think that we need more oral communicative, like more tests to assess the communicative competence or communicative ability?

R8: Yes.

INT: Rather than just pronunciation?

R8: Yeah. I think that might be, but, okay. So yeah, it would be helpful, because having students read sentences, although pronunciation wise, I guess it's a good test of it. But it also tests kind of their reading ability too, which is not really what we're testing. So, communicative competence would be nice to know. Can you, if I ask you a question, could you nicely answer it for me? But, what was I going to say? Oh, then again, because the EPT is kind of placing people also into pronunciation classes, we don't necessarily have a class that teaches like how to properly answer this question or specific vocabulary you need for this. So, it would help, but I'm not sure if we say they're not very high, they're not very competent on that, do we have a class that would help them with that? I'm not sure if we have that yet now?

INT: Mmm hmm. Okay.

R8: Don't know. [laughs]

INT: Yeah. Okay. Maybe that's all. Thank you so much.

R8: You're welcome.

(end of interview)

R9:   I have to finalize it before I get it back to the student.
INT: So you're a native speaker of English?
R9:   Correct, mmm hmm.
INT: Yeah. Then like, any other languages do you speak?
R9:   I speak Spanish as a second language.
INT: A second language. So then like, you're, what's your major right now? MATESL?
R9:   Yes.
INT: Oh, and this is the second year?
R9:   Mmm hmm.
INT: Then what was your major in undergrad?
R9:   Spanish studies, mmm hmm.
INT: Spanish studies.
R9:   I started learning Spanish in high school, and then I continued.
INT: Uh huh. So how many years have you taught English to?
R9:   Have I taught English? This is just my second year. I started teaching when I started the MATESL program.
INT: Uh huh. So did you, do you teach writing class here or?
R9:   I currently teach two classes. I teach college writing at the IEI, which is intermediate level. And then I teach a service course here. I teach ESL 506.
INT: 506. 506 is pronunciation course, right?
R9:   It's oral communication.
INT: Yeah. Oral communication.
R9:   And part of the course is focused on pronunciation.
INT: Yeah. So, so how many semesters have you taught ESL 506 or?
R9:   This is the first semester that I've taught that.
INT: First semester, yeah. So any other pronunciation courses did you teach previously?
R9:   Teaching, no.
INT: No, no. So then how many times did you participate in rating the speakers in the EPT oral section? This is the first semester?
R9:   Mmm hmm.
INT: Yeah.
R9:   I mean, prior to it, I hadn't taught pronunciation. But I went through, I had taken the phonology course with Professor Dickerson, and I went through the training with the samples and everything. So I had, I'd done some practice, but I had never actually rated before this semester.
INT: Yeah. Then what do you think the purpose of Phase 1 in the EPT oral section?
R9:   What was the purpose?
INT: Yeah.
R9:   Um, I got the impression the purpose was to sort of weed out, I guess, for lack of a better term, the people that we thought who had good pronunciation, that we would not need to go through the tests. That were very intelligible, and that we didn't think needed to even go through the interview process because they would pass.
INT: So how do you define intelligibility or that person is intelligible? How do you define that?
R9:   Me personally?

INT: Yeah, yeah. You personally and then what do you think the definition of intelligibility in the oral section of EPT, EPT oral section, yeah. Is it different?

R9:   For me it was, no, no. I mean, just like what I would consider that is how much effort I had to put into understanding them. If I could just listen like I listen to most anyone speaking English, I guess most native speakers speak English, without really trying to understand what they were saying, then to me that was very intelligible, I guess.

INT: Yeah.

R9:   And then, I guess, on the other side, if I really had to try hard either to understand what they were saying due to like individual sounds that were unclear, or their overall organization and stress and rhythm, then to me that was less intelligible or unintelligible.

INT: Yeah. So I just like to know like do you focus on every word you can understand or not? Or the general message? I mean, like can imagine the situation like when you can understand every word they uttered, but you cannot understand what they are saying.

R9:   [laughs]

INT: Or the general message, the general idea like that. So in that case, can you say that they are intelligible or not?

R9:   [laughs] I guess that's a good question. I guess they're intelligible in terms of their, the words and speaking English and the pronunciation is intelligible that you can understand that. But it's probably not coherent in the sense that it's not organized, and whatever they're saying maybe is not, they're using like direct translation, so the phrasing isn't quite right, or. Yeah. I guess that would probably be it. So in terms of intelligibility, it would be easy to understand the words but it isn't really a coherent message.

INT: Yeah. So maybe, how about like when you can understand every word they said, but their grammar is so bad. So it makes you like kind of confused like that. So that kind of situation have a kind of effect when you're rating or assessment? . . . .

R9:   It's not supposed to. According to how it is supposed to go, you're just listening to their pronunciation. We even tell them, like "I'm not listening to your grammar, so don't be nervous about that." But to be honest, it is hard to let, to let it not influence you. Because it, when you think of someone as proficient or fluent, you kind of wrap all of those things in one package. But technically you're just supposed to listen to pronunciation. So I did my best to just listen to, okay, can I understand what they're saying? It isn't necessarily that they're using good grammar or not, but the clarity of each word or each message is more or less clear. So I tried to focus on that.

INT: Yeah. So do you think that a heavy accent has a kind of an influence on intelligibility?

R9:   If it interferes, I guess, with, with understanding what they say. Like certain, I guess, qualities of accents or certain accents, I think are easier for English speakers to understand, either because they're more familiar with them or because it doesn't affect meaning as much. But if an accent, for example, causes a word to sound different, like for example, errors that often happen with "l" and "r," if it creates a different word, and so there's an error in meaning, then it does. Then it is more of a problem, and more an error in intelligibility. Whereas if a Spanish speaker says, "seet" instead of "sit" within the context, you can kind of understand what's going on, even though it sounds like a different word. So I guess the accent only kind of

factors in for me when it changes the meaning of what they're saying. And also too, if it's maybe an accent that I'm not as accustomed to. Like I've gotten used to having Korean students and Chinese students in my classes. But up until this semester, I had not had an Indian student. And so the first time when you're kind of hearing accents, and you're just getting used to it . . .

INT: Yeah, yeah .

R9: It is harder, I think, to kind of understand. Because you're not used to the way that they pronounce words.

INT: Yeah. So the kind of familiarity affect the score, assigning the score, do you think?

R9: Does, does my familiarity with it affect their score?

INT: Yeah.

R9: Um, I would try not to let it. I mean, just because I could understand what maybe the Chinese student was saying, I know it's still not correct. So I guess the familiarity factor just makes me understand or not, but it doesn't change my opinion of whether or not it's correct pronunciation.

INT: Uh huh, okay. I see. So then like in the technical terms, the suprasegmentary factors, and the segmentary factors, which one most contributes to your intelligibility, or your assessment of intelligibility?

R9: Probably suprasegmentals, when you're considering the whole message unit, not just the individual segmentals, the individual sounds. Because I think that students have a lot more errors with like linking and with rhythm and stress. And since in English we listen so much for where primary stress is, and we reduce vowels so much, when we don't hear that, I think that kind of messes you up more. Or it's harder to understand more than if like you produce a vowel a little bit off, or if you produce a "b" instead of a "v." Like the phenomena that happen on the suprasegmental a little, I think affect more the intelligibility than on the smaller level.

INT: Yeah. So then like I'm going to talk about misplacement right now. Misplacement, if you assigned like 1, a score of 1 to some test takers, so they had to go back to afternoon session. Then why did they be exempted from Phase 2 after they got referred? So the different rating, right?

R9: Mmm hmm.

INT: From your rating to the different latest rating. So what do you think, so what could be the cause or what might be the reason for the kind of misplacement, different rating, I mean.

R9: Oh, when, like if I gave a student, I sent the student to the afternoon interviews.

INT: Mmm hmm.

R9: And then they were exempted there, or something like that?

INT: Yeah, what. Yeah, what, you assign the failing score but they can, they could be exempted from Phase 2 from the later, yeah, session.

R9: Ah. It could be, in the interview in the morning, the first, the preliminary interview, it's more or less like a conversation. It's more informal. So if the student is good at interpersonal type of communication and they feel more comfortable having a conversation, they might feel more relaxed and therefore do a little better. On the opposite end, if the student gets very nervous or is shy or isn't very fluent when they have to think quickly, they could do poorly, and then I would assign a low score. Whereas in the afternoon, it's a little more structured. The student reads, so if they did poorly in the morning because they have trouble sort of having a

conversation and thinking quickly and organizing their thoughts quickly, I assigned a low score. But then in the afternoon, al they have to do is read.

INT: Yeah, read a lot.

R9: Then maybe they would make less errors because it's sort of structured for them. That could be one reason. Another thing could be, I guess, the differences among raters. There might not be a high level of rater consistency and rater reliability. Because maybe raters find different aspects of language or different features more bothersome than others. Like maybe not everyone has the same opinion as me that the suprasegmental is more important. Maybe to other people the segmental is more bothersome, so when they hear individual segments and phonemes differently, that's really bothersome, and that's wrong. And so it could be different values that the raters think are more important. Um, it could be, too, what was I going to say? Maybe the fact that it's in the morning. The student doesn't do as well in the morning.

INT: Yeah.

R9: And it's, it has nothing to do with the rater at all, but the actual student performance. That in the morning they're tired, they're not awake yet. They can't think. They didn't eat breakfast, or something. And so they come, and they do really poorly, because they just aren't awake yet. But then in the afternoon, once they've had their coffee and once they kind of had more of a chance to use English that day, they perform better, so.

INT: Yeah. Yeah. a good reason, I think.

R9: Yeah. [laughs]

INT: Yeah. [laughs] So the next question is about how do you process the Phase 1? I mean, you call the student one by one?

R9: How did I do it?

INT: Yeah, yeah. So . . . .

R9: I, yeah.

INT: You did any group format?

R9: This, so it was the first time that I had done it. I didn't know if I could, how I would organize the group discussion and like accurately like rate everyone if multiple people were talking. And since the students didn't know each other, I wasn't sure how productive a group discussion would go. So the way I did it was, I had the students sit toward the front of the room, and I called them back to me. I didn't want to sit at the front of the room, because I didn't want everyone watching. So I sat at the back and I called students back one by one, and I just talked to them for maybe 3 or 4 minutes, until I got enough of an idea of what their pronunciation was like.

INT: Yeah. So actually, uh… the intonation of students are increasing sharply. So many students have to take the EPT, so the university is encouraging to take some group format due to the time issue like that.

R9: That makes sense, yeah.

INT: Yeah, yeah. So what could be disadvantage or advantage, or like, do you like that format? And do you think it's going to work well or not? So over your opinion about good format?

R9: I see the benefit to it, that it would save time and that it, because there are so many that come through. It would be more efficient in terms of time. I'm just not sure, my doubts with it is that, like I said, I don't, I'm not sure the students would generate a very productive discussion with people they don't know at such an early

hour in the day. I can't imagine it being, you know, very talkative. And then at the same time, because it's not my class. It's not like a class of students that I know. I have to be interrupting constantly to be like, wait, what was your name, what was your name? Because it's kind of hard for the rater within 5 minutes to memorize the name, the name of 20 people in the room.

INT: Yeah, uh huh.

R9: So how do you accurately assign a score without, you know, either having them wear name tags or something. So I guess it would take a little more, maybe planning and practice. We might have to have like a practice run with students that would be willing to try it. But, I suppose it's possible. But it would take maybe some more planning than just saying you're going to interview the whole group and switching right away.

INT: Yeah, so can you, can you imagine a situation where the students can interact to each other, not just with you. So is it, so is it good for your rating? You can rate the students when they interact to each other. So is it, does it cause any problem on your rating or?

R9: Having them interact, like the way that they interact with each other versus?

INT: Yeah. You just watch what they are saying and what they are listening to, interacting things.

R9: Um, I don't think it would really affect it either way. I mean, I would still be hearing their voice, which is the important part. But if you're having, like if you set them up in pairs and they were just having a conversation with one person, you would have so much background noise that it would be hard to hear everyone. And then similarly, I guess, if you had like a whole group, you would have to deal with like the problem of turn taking and interruptions or restarting. And so it's more to get through to focus on their pronunciation.

INT: Yeah. Okay. So in the EPT testing, we just focus on pronunciation, intelligibility. So do you think that we need a more, the different test to assess their overall ability or communicative competence like that?

R9: Besides, you mean, other than pronunciation?

INT: Yeah, yeah.

R9: Uh [laughs]. That's a very good question. You mean in terms of other things I was talking about, like organization and coherence and?

INT: Yeah, yeah.

R9: And use of vocabulary, things like that.

INT: Yeah. Any developing . . .

R9: Grammatical?

INT: Yeah, grammatical, communicativability, in terms of academic context.

R9: I think that one that would definitely be important would be the communicative competence in terms of strategies that are used. Because certain students, if they, if they know that they're not getting their message across, they kind of just abandon it and move on. But it really shows something about the student and their ability to use the language if they can overcome their pronunciation weaknesses. So if they can compensate for the fact that they know they pronounce something bad by explaining it with different words. Or if they can, yeah, I guess use circumlocution to compensate for mispronunciations or misunderstandings. So I think that should definitely be taken into account or evaluated somehow. Like how the student, I guess, can use the language to make themselves understood. Um, and in terms of other things, like organization is so [laughs] hard to evaluate. I mean, even native

speakers, when they speak, are not organized. You'll probably listen to this again, and be like, what is she saying? Because it's hard to organize on the fly. So I think to evaluate students on that would be very, very hard. [laughs]

INT: Okay. I see. So maybe that's all.

(end of interview)