5-2018

A Study of ESL Students' Perceptions and Performance in One-on-One and Paired Oral Tests

Hayley Miller
St. Cloud State University

A Study of ESL Students' Perceptions and Performance in One-on-One and Paired Oral Tests

by

Hayley Miller

A Thesis

Submitted to the Graduate Faculty

Of St. Cloud State University

in Partial Fulfillment of the Requirements

for the Degree

Master of Arts in

Teaching English as a Second Language

May 2018

Committee Members:
Shawn Jarvis, Chairperson
Chookyong Kim
Semya Hakim

Abstract

The aim of this classroom-based research project was to gain insight into student perceptions and performances in two different oral assessment formats: a one-on-one interview with the teacher and a paired interaction with a classmate. The participants were 6 female and 6 male international students enrolled in high-intermediate level ESL courses in the Intensive English Center at a mid-size university. In a within-subject designed format, the students alternated the order in which they participated in the test formats, with two alternating speaking prompts. Data collected via pre-test and post-test questionnaires was analyzed in terms of seven themes of perception: *nervousness, preparedness, interest, interaction, effectiveness of format, belief in performance,* and *preference*. The results indicated that students' attitudes towards the two formats were generally positive and that there was not a significant difference in regard to themes of perception within the two formats. After performing a paired samples t-test with the average group scores from the two assessments, results revealed that the test format did not have a statistically significant effect on performance. However, as evidenced by post-test questionnaire data and student commentary analysis, it can be concluded that the two test formats are not equal interactions and should not be considered equal measurements of oral proficiency. The implication on pedagogy is that teachers should utilize the two test formats for different purposes in their assessment practices.

Keywords: oral assessment, student perceptions, assessment formats, one-on-one interviews, paired interaction

Acknowledgements

First, I thank God for leading me to this program and sustaining me throughout the journey. I express gratitude towards my thesis advisor, Dr. Shawn Jarvis, who has been instrumental in guiding this project. Her insight and experience with second language assessment was invaluable, and her confidence in my abilities was a constant encouragement. I also extend thanks to my other committee members, Dr. Chookyong Kim and Dr. Semya Hakim, for being a part of my committee, and for their insights and contributions to my research throughout this entire process. I thank every TESL faculty member at St. Cloud State with whom I had the pleasure of taking a course.

I am grateful to each of the participants in my research, who made themselves available to me amidst numerous other responsibilities, patiently participating in the assessments, and thoroughly filling out the pre-test and post-test questionnaires. Without them, this research would not have been possible.

Thank you to my mother, father, other family members, and best friends who were overwhelmingly encouraging and supportive-- you all instilled in me the strength to make such a vast undertaking as this paper possible.

Finally, I am thankful for my dreamboat Amos, whose timely appearance and constant love and support helped me in more ways than he'll ever know.

Table of Contents

Chapter Page

List of Tables

List of Figures

Chapter I: Introduction

In every classroom, student assessment plays a key role in learning and teaching. Teachers allot a significant amount of time preparing and creating instruments and observation procedures, marking, recording, and synthesizing results in informal and formal reports in their daily teaching (Cheng, Rogers, & Hu, 2004, p. 360). A major consideration in the area of language test development has been test *backwash* or *washback,* meaning the effects of testing on learning and instructional practices (Bachman & Palmer, 2010; Hughes, 2003). The test format, procedure, and content can all have a significant effect on learners, and whether the effect is harmful or helpful, it will in turn have consequences for language learning and teaching.

In the case of high-stakes testing, serious decisions are being made with the results of the test, like program or university admittance, placement within a program, graduation, and employment (Bachman & Palmer, 2010). These types of high-stakes tests are usually standardized, and their outcome has a major effect on the stakeholder's future. When it comes to low-stakes testing, like classroom assessment, the decisions made based on the results may not be as consequential. These decisions are more formative and can provide students or teachers with necessary feedback to improve learning or teaching, respectively (Bachman & Palmer, 2010). The purpose of classroom assessments is different than standardized tests in that they assess based on a more narrow and specified curriculum rather than a broad general proficiency scale.

In terms of oral assessment in language, accurately and reliably testing proficiency of second language learners has been notoriously challenging. Validity within test construct, testing format, and effectiveness of interviewers have all been questioned. Whether the oral test is taken with an interviewer or a peer, it is seen as a "co-constructed" interaction, and the participants' performances are "inextricably linked" (Brooks, 2009, p. 341-342). When

considering that the performances depend upon the other participant so closely, providing a fair format is a central component to developing oral assessments (Brown, 2003).

Few studies have been done to compare one-on-one interviews with paired interactions and on test-taker perceptions of those testing formats. Bachman & Palmer (2010) advocate for the inclusion of test-takers in the assessment development process to promote beneficial washback. This study will attempt to add to this area of research, and be guided by the principle that, "test-takers have a great deal to offer to the test researcher in making judgements about the value of the tests which they take" (Brown, 1993, as cited in Fulcher, 1996). My focus is to shed light on the perceptions and attitudes that students have about two oral testing formats - the one-on-one interview and paired interaction. This study will also consider the differences in student performance by looking at the actual scores students receive in those two formats. The research questions for this research paper are:

1. What are ESL students' perceptions of the one-on-one interview and the paired oral testing formats?
2. Is there a difference in ESL students' oral performance scores in a one-on-one interview versus a paired oral test?

The participants in this research were 12 high-intermediate level students in the Intensive English Center at St. Cloud State University. Each student participated in a one-on-one interview and a paired interaction. Using a quantitative research approach, I collected and analyzed scores of each student's performances in both of the assessment formats, their responses to closed-ended pre- and post-test questionnaires, and their responses to an open-ended post-test questionnaire. The two different types of oral language tasks were examined in terms of the effects on four scoring criteria: *Delivery, Language Use, Topic Development,* and *Interactional Competence.* Once the data was gathered, I included the results, a discussion of the implications, limitations to the study, and suggestions for future study.

Chapter II: Literature Review

Use of Language Assessments

Bachman and Palmer (2010) describe the use of language assessments as being a tool used to collect information for making decisions, and that these decisions, "will have consequences for the stakeholders, the individuals and programs in the educational and societal setting in which language assessment takes place" (p. 22). Some of the decisions made as a result of assessment may be formative, low-stakes decisions related to in-class instruction or used to improve student learning by providing feedback (Bachman & Palmer, 2010). Other decisions can have more serious, high-stake consequences for individuals like placement or acceptance into a program of study, certificate for professional employment, or passing/failing a course (Bachman & Palmer, 2010). Figure 1 (adapted from Bachman & Palmer, 2010, p. 23) shows the events that occur, beginning from the test-taker's performance and ending with the consequences of that performance. The intended use of language assessments is to accurately measure stakeholders' proficiency, and test developers and users should consider the consequences of using an assessment, and the potential decisions being made by its outcome (Bachman & Palmer, 2010).

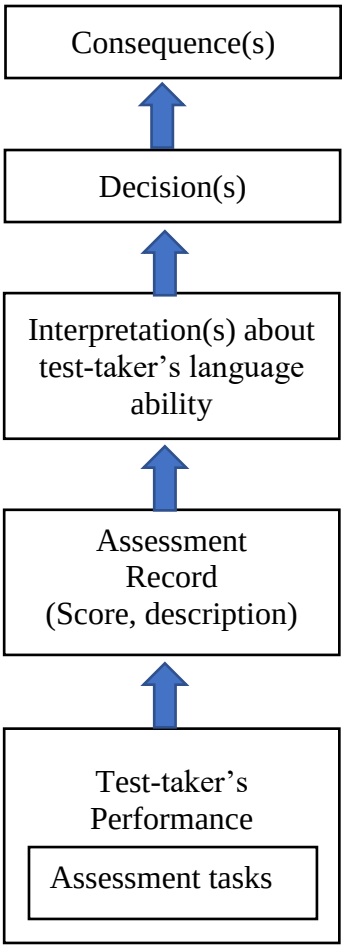

Figure 1. *Links from Test-taker's Performance to Intended Uses*

Oral Assessments

Developing a fair and useful oral assessment has been the subject of ongoing research in the field of language testing. Larry Davis (2009) describes assessment of spoken language as, "a complicated matter in which different factors interact to ultimately produce a score" (p. 367). These factors that interact are the variables that researchers have been investigating in a quest for a more reliable and valid oral assessment. Some of the variables include test format (Brooks, 2009), tasks or prompts related to performance (Bonk & Ockey, 2003; Frost et al, 2011; Fulcher, 1996; Shohamy et al, 1986), reliability of the rater(s) (Brown, 2003; McNamara, 1997; Ross, 2007), and the effects of interlocutor behavior or characteristics on performance (Davis, 2009; Lumley & O'Sullivan, 2005; Ockey, 2009; O'Sullivan, 2000).

The goal of this research is to find an oral assessment that can be more standardized and generalizable (Bachman & Palmer, 2010).

As outlined by Hughes (2003), there are three general formats for assessing oral skills: one-on-one interview with an interviewer, paired or group interaction, and voice-recorded speech. There are advantages and disadvantages that each format has, and the consequences of each vary greatly for the stakeholders.

One-on-One Interviews

A unique feature of the one-on-one interview is that it allows the interviewer to elicit specific constructs from the student that are relevant to curriculum practices. A construct, as defined by Bachman and Palmer (2010), is "an ability that provides the basis for a given assessment or assessment task and for interpreting scores derived from this task" (p. 43). Examples of constructs are knowledge of politeness markers, or knowledge of how to organize utterances to form texts (Bachman & Palmer, 2010). The purpose of the one-on-one interview in this case is to focus on eliciting a specific language construct to determine students' competence with it. The interviewer has control of the interaction, and, in this way, can manipulate the conversation in an attempt to bring out target constructs.

A growing body of research on one-on-one interviews, however, has shed light on their limitations, and questioned their appropriateness as a format to measure oral proficiency (Brooks, 2009; Brown, 2003; Kormos, 1999; McNamara, 1997; O'Sullivan, 2000; Ross, 2007; van Lier, 1989). The overall validity of the format has been questioned, specifically regarding whether or not the outcome is an accurate reflection of communicative language ability (Kormos, 1999; Ross, 2007; van Lier, 1989). The nature of the one-on-one interview format represents an imbalance of power between the interviewer and test-taker. Kormos (1999), concludes that the interviews are "unequal social encounters" that do not resemble authentic conversational interactions (p. 164). This is seen as a drawback when considering

conversational ability to be, “an appropriate vehicle for the all-around display of speaking ability in context” (van Lier; 1989, p. 489). The unequal roles that the tester and test-taker play in the interaction is seen as problematic, as well as the emphasis on elicitation rather than conversation (van Lier, 1989).

The interaction as elicitation and not true conversational discourse is another limitation discussed by Young and Milanovic (1992) and van Lier (1989) using the conceptual framework of dyadic interaction proposed by Jones and Gerard (1967). Jones and Gerard (1967) outline the model of dyadic interaction as being the way that people, “behave in each other’s presence” (p. 505). When people interact, they bring certain goals to the situation that motivate their conversation (Jones & Gerard, 1967). During the one-on-one interview, the interviewer and interviewee have different goals. The interviewer’s goal is elicitation, and because they hold more social power and control, the interviewee’s goal is contingent upon the interviewer (Jones & Gerard, 1967; van Lier, 1989; Young & Milanovic, 1992).

This asymmetrical contingency is defined by Jones & Gerard (1967) as a class of interaction wherein, “the responses of one person are largely determined by self-produced stimuli or plans, whereas the responses of the other are largely determined by social stimuli produced by the first” (p.509). The one-on-one interview, being characterized by asymmetrical contingency, may become less of a conversation and more of an interrogation (van Lier, 1989). In this environment, the interviewee may feel threatened by the interaction (Bachman & Palmer, 2010). Young and Milanovic (1992) refer to this kind of conversational dominance as being the tendency for one person to control the interaction by means of initiating and ending topics, holding the floor, and controlling the other participant’s access to the conversation by interrupting or asking questions.

Another issue that affects reliability of the one-on-one interview is the difference in discourse style and behavior of the interviewers, which Ross (2007) calls "inter-interviewer variation" (p. 2019). Ross explains, "A key assumption of the OPI [oral proficiency interview] is that interviewers and raters are completely interchangeable – a candidate's performance in an interview and the rating of that performance are presumed to be independent of the idiosyncrasies and interaction style of the interviewer" (p. 2019). This behavior variation can be expressed through differences in facilitative accommodation (Ross & Berwick, 1992; Ross, 2007), sharing common interest with the interviewee (Ross, 2007), and extent of transitional talk (Ross, 2007). Ross (2007) describes the potential for the interviewer to skew the interviewees "chances of comprehension and uptake" by varying in amount of discourse that expresses shared interest with the interviewee and transitions one task to the other (p. 2022). If an interviewer does not utilize shared interest with the interviewee to link the topics of conversation, but moves quickly from one topic to the next, it is less likely that the interviewee will follow the frame of conversation. On other cases, Lumley and Brown (as cited in Brooks, 2009) found that the interviewer simplified their language to aid the interviewee. Ross (2007) found that interviewers tended to accommodate by "scaffolding the interaction" (p. 2017). These features of conversation building and facilitating may change from one interviewer to the next, varying the performance of interviewees, and, ultimately affecting raters' perceptions of interviewee ability (Brown, 2003). The "lack of standardization across interviews" decreases the overall reliability of the format and has potential unfairness for participants (Brown, 2003, p. 1).

The feature of interviewer gender and its effect on performance has also been the subject of studies in language testing (Lumley & O'Sullivan, 2005; O'Sullivan, 2000; O'Loughlin, 2002). Early studies were done by Locke (as cited in O'Sullivan, 2000) and Porter (as cited in O'Sullivan, 2000) on Arab and Algerian students, and found that, when the

interviewer was a male, all of the interviewees achieved a higher score. Porter and Shen (as cited in O'Sullivan, 2000), however, went on to study students with mixed nationalities, and found that they achieved higher scores when interviewed by a woman. These studies show that, indeed, there is a correlation between gender of interviewer and the performance of the interviewee. The gender correlation, however, may vary depending on the cultural background of the interviewee. Barry O'Sullivan (2000) sought to examine the effects of the interviewer's gender on performance and compared the one-on-one performances of 12 students when they interviewed with a female examiner and when they interviewed with a male examiner. After evaluating the performances, O'Sullivan concluded that the students performed better when the examiner was a female, regardless of the gender of the test-taker. However, in a study done in 2002, Kieran O'Loughlin found that the gender of the interviewer made no significant difference. He studied the outcome of eight female and eight male students' interviews with a female interviewer and then with a male interviewer and found no clear association between interviewer gender and interviewee performance (O'Loughlin, 2002). Lumley and O'Sullivan (2005) more recently studied the effects of gender on performance via a tape-mediated speaking assessment and found little correlation between gender and performance outcome. They concluded that, if gender plays a role in performance, it was not consistently the case and cannot reliably be predicted (Lumley & O'Sullivan, 2005).

Paired Interaction

Due to the limitations of the one-on-one interview, paired interaction emerged as a newer, alternative testing format (Leaper & Riazi, 2013). Growing investigation on group and paired interaction has been exploring the various advantages and disadvantages of the format (Bonk & Ockey, 2003; Brooks, 2009; Davis, 2009; McNamara, 1997; Nitta & Nakatsuhara, 2014; Ockey, 2009; Shohamy et al, 1986; Taylor & Wigglesworth, 2009; Van Moere, 2006).

The focus of this research has been on discourse of the interaction (Brooks, 2009; Bonk & Ockey, 2003; McNamara, 1997; Taylor & Wigglesworth, 2009) and various features of group members, and the effects those features have on the other's performance (Brooks, 2009; Davis, 2009; Ockey, 2009; Van Moere, 2006).

One advantage of the paired interaction is that it more accurately reflects the pair and group work taking place in the language classroom (Bonk & Ockey, 2003). This type of test format mimics authentic conversation, which occurs in the language classroom, thus has a potential positive washback in teaching and learning (Ockey, 2009). Since the participants in the assessment interaction have similar purposes, their conversation reflects mutual contingency rather than the asymmetrical contingency in the one-on-one interview (Jones & Gerard, 1967). Jones and Gerard (1967) describe interactions in the mutual contingency class as requiring "that a plan govern the responses of each actor, but the plan becomes continually recast in the light of the other's responses" (p. 511). The participants are considered social equals in the paired interaction because they are peers, and so their goal-orientation and reactiveness to each other in the conversation are symmetrical (Kormos, 1999). Van Lier (1989) suggests that assessments in peer groups reduce asymmetry in conversation. According to Leaper and Riazi (2013), the shift from one-on-one interviews to paired interaction reflects the "move from conceiving of speaking ability as represented by the linguistic features of an individual's spoken words to one of interactive communication" (p. 177).

Another advantage to this format is that it is received positively by test-takers (Fulcher, 1996; Shohamy et al, 1986; Van Moere, 2006). Shohamy, Reves, and Bejarano (1986) experimented with four different oral assessment tasks and formats on a group of 103 students. One of the formats included was a group discussion, and while it was rated less favorably among students compared to the one-on-one interview and role-play, the students

generally favored it as a part of the overall group of assessments. Glenn Fulcher (1996) studied 47 students' performances and perceptions of three different tasks in an oral assessment. Two of the tasks were one-on-one interviews and the other task was a group discussion. The results indicated that well over 50% of the students found the group discussion to be more enjoyable than the one-on-one interviews. According to two of the participants, the group interaction format reduced anxiety. More recently, Alistair Van Moere (2006) examined 113 students' performances and perceptions in a group oral discussion and found that the students gave positive reactions to the test format.

There are, however, some potential disadvantages in the paired interaction due to the dependence of performance on a partner (Brooks, 2009; Davis, 2009; Ockey, 2009; Van Moere, 2006). Within the paired oral test, the members participating in the interaction each contribute to the performance, and therefore their performances are inextricably connected (Brooks, 2009). This means that a co-constructed test performance could be affected by a variety of other group member variables such as gender, proficiency level, and personality (O'Sullivan, 2000).

A study done by Larry Davis (2009) explored the possibility that the partner's proficiency level could affect performance. He compared two test performances: one in which another student had a similar level of proficiency, and one in which another student had a higher or lower proficiency level. Using the results, Davis concluded that the proficiency level of the partner did not have an observable effect on performance. He did, however, find that the students with the lower proficiency performed better when paired with a higher-level partner. Natkatsuhara (as cited in Davis, 2009) concluded similarly that the proficiency level had little effect on the overall score. In contrast to Davis (2009) though, it was found that higher-level test-takers performed slightly better when paired with lower-level test-takers. An earlier study by Iwashita (as cited in Davis, 2009) found that both high-level

and low-level students achieved a better score when paired with a partner of a different proficiency level.

When considering the personality of individual test-takers, there are a lot of features that can affect the other test-taker's performance, for example level of extroversion, shyness, and/or dominance (Ockey 2009; Van Moere, 2006). Gary J. Ockey (2009) focused on the assertiveness of students in correlation with group test performance by highlighting the effect on test performance of having assertive versus non-assertive group members. In his study, groups of four students were assigned based on their levels of assertiveness; one group had all assertive personalities, one group had all non-assertive personalities, one group had a majority of assertive personalities, and the last group had a majority of non-assertive personalities. The participants' level of assertiveness was measured based off of their results of their revised NEO Personality Inventory. The results showed that the only students whose performances were affected were the assertive students: they performed better when in a group of non-assertive test takers and worse when in a group of only assertive test takers. The non-assertive test takers appeared to be unaffected by the different personality testing environments.

A Study Comparing Two Assessment Formats

Lindsay Brooks (2009) followed Vygotsky's framework of sociocultural theory of mind (SCT) to compare quantitative and qualitative differences in performance when the same test-taker interacts one-on-one with an interviewer and when they interact in pairs. The position of her study was that the nature of an interaction is co-construction between the involved participants. The performance of one person involved in the exchange is dependent upon the other person involved, and their action is inseparable. As the interaction takes place, the participants are building an oral performance jointly. She sought to explore how the test-

takers' performances differ in each format and what the features of interaction are in each format.

In her study, test-takers took the test in each format with comparable discussion prompts. Each test was independently evaluated by two raters with identical holistic rating scales. The results from the students' scores revealed that students performed better in the paired format than the one-on-one interview. The pair interaction format produced a greater range of features of interaction. The study found that the one-on-one interview reflected asymmetrical discourse, and a majority of the features of interaction were questions posed by the interviewer. Brooks concluded that the paired interaction represented a more co-constructive, collaborative dialogue.

Though Brooks' conclusion is significant to oral assessment development, there is a shortage of research focused on comparing performance in one-on-one interviews with paired interaction. The potential advantages of group or paired oral assessment over one-on-one interviews suggests that more research is needed (Van Moere, 2006). Van Lier (1986) also describes a lack of research investigating task-based assessments with peers as a feasible alternative to one-on-one interviews.

Oral Assessment and Student Perceptions

In order to investigate the creation of a more fair and reliable oral assessment, test-takers' involvement is necessary (Bachman & Palmer, 2010). Bachman and Palmer (2010) emphasize the test-taker's role in assessment development:

> One way to promote the potential for positive consequences of assessment use is through involving test takers in the development of the assessment, as well as collecting information from them about their perceptions of the assessment and the assessment tasks. If test takers are involved in this way, we would hypothesize that the assessment tasks are more likely to be perceived as authentic, and that test takers will have a more positive perception of the assessment, be more highly motivated, and probably perform better (p. 107).

When it comes to test-taker perceptions of paired or group oral assessments, some research has shown that they have positive reception toward the format (Fulcher, 1996; Shohamy et al, 1986; Van Moere, 2006). Because of this favorable outlook, more research is necessary to expand and validate the positive perception (Van Moere, 2006).

Chapter III: Methods

Research Questions

In light of this position warranting test-taker involvement in assessment development, and the shortage of research investigating one-on-one interviews versus paired interactions as assessment formats, I proposed an investigation seeking to explore ESL students' perceptions of two different oral testing formats, individual interview and paired oral test, and explored a connection between students' perceived performance and actual performance in both testing formats. The research questions for this research paper are:

1. What are ESL students' perceptions of the one-on-one interview and the paired oral testing formats?
2. Is there a difference in ESL students' oral performance scores in a one-on-one interview versus a paired oral test?

Participants

Test-takers. The participants in this study were 12 high-intermediate ESL students attending a mid-size public university in the United States. The students were enrolled in high-intermediate courses in the Intensive English Center (IEC) on campus where they participate in 23 hours of English instruction per week. Placement into the high-intermediate IEC courses was based on test scores from the paper-based Cambridge Michigan Language Assessments (CaMLA) English Placement Test (EPT) and an essay writing test. The students took the EPT and writing test before entering their first semester of courses. Those enrolled in the courses expect to matriculate into an undergraduate program on the condition of passing the level in accordance with IEC and university standards.

The participants that comprised the course were 6 female and 6 male international students from various countries and language backgrounds, (See Table 1 for an overview of

participants). Prior to their entrance into the IEC, each of them had different amounts of English instruction, exposure, and experience. Though they came into the IEC with various backgrounds, they have all tested into the IEC high-intermediate level, reflecting their similar level of English competency. To ensure anonymity, each of the test takers was assigned a number to represent their identity throughout this paper.

Table 1. *Overview of Participants*

	High-intermediate students (N= 12)
Gender	
Male	6
Female	6
Native Language	
Mandarin Chinese	7
French	2
Arabic	2
Mongolian	1

Raters. Two teaching assistants in the TESOL program scored the participant performances using voice recordings and the scoring scale provided. The raters had time to review the scoring scale and were briefed on the testing formats.

Materials

Pre-test background survey. Prior to the oral tests, the students received a background survey with seven closed-ended items regarding some personal details, preferred language activities, and perceptions about speaking tests. See Appendix A for the full background survey.

Speaking prompts. There were two prompts used in the oral tests: one was “What Fear Can Teach Us” and the other “Fear and Media”. Each student encountered one of the prompts in the first test, and then the other prompt in the second test. The prompt task design was adapted from Song (2014) in terms of format, instructions, and length. Every effort was

taken to make sure that the prompts were comparable. Each of the prompts had a general topic that students were familiar with during the course of the semester. The specific layout and contents of the prompts can be seen in Appendix B.

Post-test questionnaires. After each test, the students took a post-test questionnaire (PTQ) regarding their experience during the tests and thoughts after it (See Appendix C). After the first test, the questionnaire consisted of fifteen items which the students rated on a ten-point Likert scale. After the second test, the questionnaire was identical to the first one, but with an additional second part. The second part consisted of four closed-ended items and two open-ended items. The closed-ended items asked the students to compare the two test formats, and open-ended items elicited information about their perceptions of the two testing formats. The questionnaire items were adapted from Song (2014).

Scoring scale. The scoring scale used was replicated from Song (2014). Song developed the scale taking into account models of language ability and oral ability from Bachman and Palmer (2010) and Fulcher (1996), the speaking construct from the TOEFL iBT, and communication skills construct from Ockey (2009). See Song (2014) for a full description of scoring scale development. There were four scoring categories: delivery, language use, topic development, and interactional competence (see Appendix D). In each category, there were performance descriptors that correlated with a 0-4-point value. From the scale, the total possible score was 16 points.

General Design

Within-subject design. This study was a within-subject design to counterbalance the students with the order of the testing formats and prompts. There were two sessions of testing, and each student took both tests and questionnaires by the end of the second session. In the first session, the 12 students were instructed to select a partner to participate with in the paired test. Then the students were divided into two groups (Group A and B), making sure

that each self-selected pair was in the same group. Group A took the individual format test with the "What Can Fear Teach Us" prompt, and Group B took the paired format test also with the "What Can Fear Teach Us" prompt. In the second session, Group B took the individual format test with the "Fear and Media" prompt, and Group A took the paired format test also with the "Fear and Media" prompt. See Table 2 for the complete design.

Table 2. *Within-Subject Design*

	Individual Format	Paired Format
Session 1	Group A (N-1, 2, 3, 4, 5, 6) *What Can Fear Teach Us* Prompt	Group B (N-7, 8, 9, 10, 11, 12) *What Can Fear Teach Us* Prompt
Session 2	Group B (N-7, 8, 9, 10, 11, 12) *Fear and Media* Prompt	Group A (N-1, 2, 3, 4, 5, 6) *Fear and Media* Prompt

Procedures

Recruitment and consent forms. The students were enrolled in the high-intermediate Listening and Note-Taking course when participating in this study. I, the researcher and their classroom teacher, recruited all students in the class to take part in this study. The participating students then received a consent form outlining the study and the details of the procedure. The study was explained to them, and detailed instructions were given before the tests took place. When the consent forms were obtained, the data collection began.

Pre-test background survey. During the first session, the students all receive a background survey. They were given instructions to choose answers for the close-ended

items regarding some personal information, their preferred language activities, and preferences for speaking test formats.

First oral tests. After the background questionnaire, the students were divided into two groups of six (Group A and B), keeping the self-selected pairs in the same group. Group A were students who took the oral test in the individual format, and Group B were students who took the paired oral test. Both of the groups used "What Can Feat Teach Us" prompt for the first test.

In the individual format, the student sat across from the interviewer. The student received the prompt to read and had one minute to think about how they would respond. After one minute, the interviewer said, "You can begin speaking." Every test was voice-recorded. When five minutes was completed, the student was told to stop.

In the paired format, the students sat face-to-face. I sat outside the group as the observer. The students received the prompt to read and had one minute to think about how they should respond. After one minute, the observer said, "You can begin speaking." Every test was voice-recorded. When five minutes was finished, the students were told to stop.

First post-test questionnaire. After the first tests were completed, all of the students took the first post-test questionnaire. The questionnaire was done immediately after completing the tests to gauge the students' perceptions of their experiences and performances.

Second oral tests. During the second session, Group B began with individual format tests using the "Fear and Media" prompt. Group A followed with paired tests using the "Fear and Media" prompt as well. The procedure for the individual and paired tests was the same as the first test procedure.

Second post-test questionnaire. After completing the second oral test, the students all completed a second post-test questionnaire. This questionnaire was identical to the first

post-test questionnaire, consisting of fifteen Likert-scale items. Following the fifteen closed-items, was a second part of the questionnaire (only done after the students completed both test formats).The questionnaire was done immediately after completing the tests to gauge the students' perceptions of their experiences and performances.

Rating procedures. Two IEC teachers used the voice-recordings and scoring scale to rate all of the tests. Before they began scoring, they were trained on the scoring scale by reviewing each of the categories, the correlating descriptions, and meaning behind the constructs. The two raters listened to the voice-recordings of the tests and scored each student. After scoring all of the tests, the scores were compiled as data. Each student had a total of four test scores from the raters (two scores for the individual test format and two for the paired test format). The average score from the two raters for each test format was taken, so each student ultimately had two scores- one from each testing format.

Data Analysis

Once all of the data was gathered from the data collection instruments, the information was compiled into tables to represent individual student data and whole group data. In order to answer the first research question, data from the background survey, the post-test questionnaires, and the post-test questionnaire part 2 was taken and analyzed in terms of descriptive statistics. The goal was to compare students' perceptions before the assessments to their perceptions immediately after participating in each test to find patterns that would indicate feelings about each test format. In order to answer the second research question, scores from the assessments were compiled and descriptive statistics were used to compare individual and group scores in the one-on-one test format with the paired test format. Then, a paired sample t-test was done to discover whether or not the two test formats yielded statistically significant differences in scores.

Chapter IV: Results

Research Question 1: Student Perceptions of Test Format

The results of research question one are discussed in terms of seven themes of perception that emerged from the data: *nervousness, preparedness, interest, interaction, effectiveness of format, belief in performance,* and *preference.* The results indicated that students' attitudes towards the two formats were generally positive and that there was not a significant difference in regard to perceptions of the two formats. Overall, the students reported low nervousness, that they were prepared, interested, had ease in the interactions, believed the tests to effectively measure their English ability, and were confident in their performance in both the one-on-one and paired formats. Students indicated a division in their preference of the two formats. The following sections in this chapter are detailed descriptions of the results, relating the themes with the data collection instruments: 1) the pre-test background survey; 2) the two post-test questionnaires; 3) the post-test questionnaire part 2 and 4) the post-test questionnaire part 2 commentary. As each of the themes of perception vary in number of data collection instruments used, the introduction to each theme outlines which ones were used.

Nervousness. Evidence from the background survey, the post-test questionnaires, the post-test questionnaire, part 2, and the post-test questionnaire, part 2 commentary reveal that students generally did not feel nervous before and during both the one-on-one and paired test formats, with the one-on-one test format being slightly more nerve-racking. Detailed results from the data collecting instruments follow.

Background survey. The results from item 5 reveal differences in attitudes toward test formats. Based on question 5, six (50%) of the twelve participants reported more nervousness in a one-on-one format with the teacher, two (~17%) reported being more nervousness in a paired setting and four (~33%) reported an absence of nervousness in either

format (see Table 3). The findings suggest the one-on-one format was the most nerve-racking.

Table 3. *Background survey item 5 results* (N = 12)

Item	One-on-one	Paired	Other answers
Q5 Which test makes you more nervous?	6 (50%)	2 (~17%)	Neither – 4 (~33%)

Post-test questionnaires. The results for three items within the theme of nervousness (questions 1, 3, and 4) suggest students were slightly more nervous before and during the one-on-one test, and that they felt slightly less nervous in the paired format (see Table 4 for detailed results). In contrast to their initial nervousness about test formats in the background survey, students' mean scores on questions 1, 3 and 4 indicate a low level of nervousness before and during both test formats, with the paired test slightly lower. The results also suggest that during actual test conditions, students felt generally comfortable in a one-on-one and paired situation, with only slightly less nervousness with another student.

Table 4. *Nervousness perceptions in post-test questionnaires* (N = 12)

Category	Item	One-on-One Mean (SD)	Paired Mean (SD)	Difference
Nervousness	Q1	4.17 (2.89)	3.75 (2.61)	0.42
	Q3	4.67 (3.73)	3.50 (2.67)	1.17
	Q4	7.50 (2.43)	8.08 (1.88)	0.58

Note. 1: Strongly disagree – 10: Strongly agree

Post-test questionnaire, part 2. The results confirm that the one-on-one format did produce more nervousness than the paired format, although it was lower than students had initially anticipated in the background survey: approximately 33% reported being more nervous in the one-on-one format (see Table 5). Surprisingly, the same number (~17%) as in the background survey, felt more nervous in the paired format. And, in contrast to the

background survey, after actual test conditions, the majority (~42%) of students reported not feeling nervous in either format.

Table 5. *Post-test questionnaire, part 2 item 2 results* (N = 12)

Item	One-on-one	Paired	Other answers
Q2 Which test made you more nervous?	4 (~33%)	2 (~17%)	Both – 1 (~8%) Neither – 5 (~42%)

Post-test questionnaire, part 2 commentary. This data confirms that the paired format was less nerve-racking. Approximately 33% of students commented on having low nervousness during the paired test, and higher nervousness during the one-on-one test- the same number (~33%) as in the PTQ 2 question 2 (see Table 6 for comments). In contrast, one student commented on being comfortable in the one-on-one test. These comments confirm the overall results from the previous data reports on nervousness, which reveal that students generally felt the one-on-one test was more nerve-racking.

Table 6. *Nervousness in post-test questionnaire, part 2 commentary* (N = 12)

Category	One-on-One	Paired
Nervousness	*I'm nervous, but comfortable…* (S1)	*Also nervous, but it is interesting.* (S1)
	…I was really nervous. (S3)	*…I felt comfortable…* (S3)
	I think is very comfortable when I talk to my teacher. I don't feel nervous. (S5)	*In this test I felt more comfortable…* (S8)
	I felt nervous a little bit. (S8)	*I feel relax, just like a normal conversation. I don't feel so nervous.* (S10)
	I feel super nervous about one-on-one test. (S10)	*…feel relax.* (S12)

Note. S = Student.

Preparedness. Evidence from the post-test questionnaires, the post-test questionnaire, part 2, and the post-test questionnaire, part 2 commentary indicate that students generally felt prepared for both the one-on-one and paired test formats. Detailed results from the data collection instruments follow.

Post-test questionnaires. The results reveal that students generally felt prepared for both test formats, but slightly more prepared for the paired format. Table 7 outlines the results from item two.

Table 7. *Preparedness perceptions in post-test questionnaires* (N = 12)

Category	Item	One-on-One Mean (SD)	Paired Mean (SD)	Difference
Preparedness	Q2	6.92 (1.68)	7.42 (1.93)	0.50

Note. 1: Strongly disagree – 10: Strongly agree

Post-test questionnaire, part 2. In contrast to the post-test questionnaires, these results indicate that a majority of students felt slightly more prepared for the one-one-one test (see Table 8). The data shows that 50% of students felt more prepared for the one-on-one test versus the other approximately 42% of students who felt more prepared for the paired test.

Table 8. *Post-test questionnaire, part 2 item 4 results* (N = 12)

Item	One-on-one	Paired	Other answers
Q4 Which test do you were more prepared for?	6 (50%)	5 (~42%)	Neither – 1 (~8%)

Post-test questionnaire, part 2 commentary. In the comments, one student (student 3) reported not being prepared for the topic of the prompt in either test (see Table 9). Another student (student 10) commented on the difficulty in preparing for the paired test because it was difficult to predict what the other person would say. These comments reflect the complexity and range of what it means for a student to be prepared.

Table 9. *Preparedness in post-test questionnaire, part 2 commentary* (N = 12)

Category	One-on-One	Paired
Preparedness	*I didn't choose the topic I know well.* (S3)	*I was not prepared for this topic…* (S3) *…it's hard to prepare for it because you don't know what others said.* (S10)

Note. S = Student.

Interest. Overall, evidence from the post-test questionnaires, the post-test questionnaire, part 2, and the post-test questionnaire, part 2 commentary revealed that students are interested in both test formats, with a slight increase in interest in the one-on-one format. Detailed results from the data collection instruments follow.

Post-test questionnaires. The findings from question 5 showed that students found the paired test to be slightly more interesting but generally both formats yielded positive perceptions of interest (see Table 10). The mean scores differed by only .66, yet the standard deviation for the paired mean was only 1.00. This indicates that, generally, students deviated from the 9.08 mean by only 1 point. The standard deviation from the one-on-one mean of 8.42 was 2.27, making the degree of separation from the 8.42 mean a lot wider.

Table 10. *Interest perceptions in post-test questionnaires* (N = 12)

Category	Item	One-on-One Mean (SD)	Paired Mean (SD)	Difference
Interest	Q5	8.42 (2.27)	9.08 (1.00)	0.66

Note. 1: Strongly disagree – 10: Strongly agree

Post-test questionnaire, part 2. Contrary to the post-test questionnaires, these results from item 3 show that students were more interested in the one-on-one format. 50% of students marked the one-on-one format to be more interesting, approximately 33% marked the paired format, and approximately 17% marked that both test formats were interesting (see Table 11).

Table 11. *Post-test questionnaire, part 2 item 3 results* (N = 12)

Item	One-on-one	Paired	Other answers
Q3 Which test was more interesting?	6 (50%)	4 (~33%)	Both – 2 (~17%)

Post-test questionnaire, part 2 commentary. Four total students commented positively about the paired format being interesting, and one student commented positively about both the one-on-one format and the paired format as being interesting (see Table 12). One student commented that the paired format was not interesting because students have the same opinion as each other. Another student reported more interest in the one-on-one format. When comparing individual students' comments to the answers they indicated in the post-test questionnaire, part 2 item 3, it shows that the students who indicated being more interested in the paired format also commented about being interested in it in the open-ended commentary. In comparison, only one of the students who indicated more interest in the one-on-one format commented about it in the open-ended commentary. Students more interested in the one-on-one format tended not to comment about it.

Table 12. *Interest in post-test questionnaire, part 2 commentary* (N = 12)

Category	One-on-One	Paired
Interest	*…it is interesting to have a conversation with the teacher.* (S10)	*…it is interesting.* (S1) *I think it is very interesting…* (S2) *…we give some interesting informations.* (S3) *…we always have same point. This is not interesting.* (S9) *…interesting.* (S10)

Note. S = Student.

Interaction. Evidence from the post-test questionnaires, and the post-test questionnaires, part 2 commentary shows that students generally indicated positive perceptions of the interactions and ease of speaking in both of the formats. There were reports of accommodation in the one-on-one format and difficulty of interaction in the paired format, indicating the one-on-one and paired test formats are different interactions. Detailed results from the data collection instruments follow.

Post-test questionnaires. Items six through twelve of the post-test questionnaires represented views on interaction, and the mean score differences in the formats were within one point of each other (see Table 13). The only pattern recognized was that students reported that in the paired format it was slightly easier taking turns (Q6), they could speak when they wanted (Q8), they could explain their ideas well (Q9), and they could ask questions more easily (Q10).

Table 13. *Interaction perceptions in post-test questionnaires* (N = 12)

Category	Item	One-on-One Mean (SD)	Paired Mean (SD)	Difference
Interaction	Q6	7.50 (2.24)	7.83 (2.04)	0.33
	Q7	4.25 (3.65)	3.83 (3.56)	0.42
	Q8	7.33 (2.46)	7.58 (2.50)	0.04
	Q9	7.00 (2.17)	7.17 (2.79)	0.62
	Q10	6.58 (2.61)	7.42 (1.62)	0.99
	Q11	2.83 (3.10)	3.50 (3.48)	0.38
	Q12	7.67 (1.97)	7.25 (2.53)	0.42

Note. 1: Strongly disagree – 10: Strongly agree

Post-test questionnaire, part 2 commentary. Students commented about various aspects of interaction in both formats, as outlined in Table 14. Two students commented positively about the teacher accommodating them in the one-on-one test when they could not explain themselves. This supports item number 12 in the post-test questionnaires, in which students indicated they found it slightly easier to say what they wanted to say in the one-on-one format (see Table 13). Two other students described some difficult aspects of the interaction in the paired format, i.e. not being able to understand the other student's speech,

or the inability to clearly relate ideas to the other, resulting in frequent pausing. These comments slightly contradict item 8 in the post-test questionnaires, in which students indicated more ease in explaining ideas in the paired test format.

Table 14. *Interaction in post-test questionnaire, part 2 commentary* (N = 12)

Category	One-on-One	Paired
Interaction	*…when I cannot talk, or my brain stopped, the teacher asked me, then I can talk.* (S4)	*…talk with another student have a little difficult because we can't describe clearly what we want to say, so we pause during the conversation.* (S5)
	It can help me explain my ideas well. (S10)	*We can share our ideas for each other.* (S1)
	When I forget what I should say, the teacher will help me. (S12)	*Sometimes I don't understand partner and another student's speech. I can't say "I don't understand", just listen to continue.* (S6)
		…helped me in knowing other ideas from someone else. (S7)

Note. S = Student.

Effectiveness of format. This category elicited student feelings concerning whether or not the formats were effective in showing their English skills using the post-test questionnaires, and the post-test questionnaire, part 2 commentary. Generally, students indicated that both test formats were effective, with the one-on-one format being slightly more effective. Detailed results of the data collection instruments follow.

Post-test questionnaires. According to the results, students felt that both formats allowed them to show their English ability (see Table 15). However, the higher mean score from the one-on-one post-test questionnaire reveals that students view the format as slightly better for showing their English skills.

Table 15. *Effectiveness of format perceptions in post-test questionnaires* (N = 12)

Category	Item	One-on-One Mean (SD)	Paired Mean (SD)	Difference
Effectiveness of Format	Q13	9.17 (1.11)	8.42 (1.83)	0.75

Note. 1: Strongly disagree – 10: Strongly agree

Post-test questionnaire, part 2 commentary. The comments elicited from students confirmed that students felt that the one-on-one test format was more effective for showing their English skills. Table 16 presents the comments.

Table 16. *Effectiveness of format in post-test questionnaire, part 2 commentary* (N = 12)

Category	One-on-One	Paired
Effectiveness of Format	*I think it is better to show the students' real level of English skill. And it can let teacher know their shortcoming, then help them improve their abilities.* (S2) *I think it's best way to know students' English level.* (S6)	No comments.

Note. S = Student

Belief in performance. Results from the background survey, the post-test questionnaires, and the post-test questionnaire, part 2 reveal that students were generally more confident about their performance in the one-on-one format than the paired format. Detailed results from the data collection instruments follow. Further results for the belief in performance results in correlation with actual performance scores are discussed in the results for research question 2 section.

Background survey. Four students (~33%) initially predicted that they could perform better in the one-on-one test format and two (~17%) reported that they could perform better in the paired test format. Four other students (~33%) were not sure about their performance in either format, and the last two students (~17%) indicated that the format would make no difference to their performance (see Table 17).

Table 17. *Background survey item 6 results* (N = 12)

Item	One-on-one	Paired	Other answers
Q6 Which test do you think you can do better on?	4 (~33%)	2 (~17%)	I don't know – 4 (~33%) No difference – 2 (~17%)

Post-test questionnaires. It appears that students believed positively in their performances in both testing formats (see Table 18). Item 15 directly asked about belief in performance, and there was a slight increase in overall student belief in the paired test performance. This slightly contradicts the background survey results, in which students believed in their one-on-one performance.

Table 18. *Belief in performance perceptions in post-test questionnaires* (N = 12)

Category	Item	One-on-One Mean (SD)	Paired Mean (SD)	Difference
Belief in Performance	Q14	7.50 (1.98)	7.25 (1.96)	0.25
	Q15	6.58 (2.47)	7.08 (2.47)	0.50

Note. 1: Strongly disagree – 10: Strongly agree

Post-test questionnaire, part 2. Seven students indicated their belief that they performed better with the teacher, and the other five believed that they performed better with another student (see Table 19). This shows that, like the background survey, more students believe in their one-on-one test performance over their paired test performance.

Table 19. *Post-test questionnaire, part 2 item 1 results* (N = 12)

Item	One-on-one	Paired	Other answers
Q1 In which test do you think you got a higher score?	7 (~58%)	5 (~42%)	-

Preference. Results from the post-test questionnaire, part 2, and the post-test questionnaire, part 2 commentary indicate that students are divided in their preference of the one-on-one format versus the paired format. Detailed results from the data collection follow.

Post-test questionnaire, part 2. Five students (~42%) reported that they preferred the one-on-one format, five (42%) preferred the paired format, and two students (16%) preferred both formats equally (see Table 20).

Table 20. *Post-test questionnaire, part 2 item 5 results* (N = 12)

Item	One-on-one	Paired	Other answers
Q5 Which test style do you prefer?	5 (~42%)	5 (~42%)	Both – 2 (~16%)

Post-test questionnaire, part 2 commentary. One student went on to comment about the paired format not being their preferred format (see Table 21). However, that individual student indicated in the post-test questionnaire, part 2 that the paired test was the test format they preferred, thus contradicting themselves.

Table 21. *Preference in post-test questionnaire, part 2 commentary* (N = 12)

Category	One-on-One	Paired
Preference	No comments.	*I think it was not preferred…* (S3)

Note. S = Student

Research Question 2: Student Performance Scores

The mean test scores from the twelve students' one-on-one and paired tests indicated that the test format did not have a significant effect on performance. Table 22 presents descriptive statistics of the two test format results. The paired samples t-test ($t = .382$ [11], $p < .710$) showed that the difference between the two group means was not statistically significant. A count of the students' scores (see Table 23) reveals that five students performed better in the paired format, five other students performed better in the one-on-one format, and two other students' scores remained unchanged between formats. The order of testing (i.e. whether the students tested first in pairs or one-on-one) did not seem to have an

effect on the scores. Group A had a slightly higher mean score than Group B in both of the formats.

Table 22. *Descriptive statistics for test-takers' performances in two formats* (N = 12)

Test format	Mean	SD	Median	Min.	Max.
One-on-one	9	1.70	9.5	6.5	11.5
Paired	8.83	1.91	8.75	6.5	12.5

Note. Score out of 16 points.

Table 23 presents the individual outcomes of the students, their belief in outcome, and their actual outcome. Individually, there were not large discrepancies between performance outcomes in either format. The greatest change was from student nine who received a ten point score in the one-on-one test and then a seven point score in the paired test. Of the seven students who believed in having a better outcome in the one-on-one test, five of those students did receive a higher score in that format. Five students believed they received a higher score in the paired test and three of those students did receive a higher score in that format. Of the twelve total students, eight (~67%) confirmed their belief and the other four (~33%) did not confirm their belief, but either remained the same or were inaccurate.

Table 23. *Test-takers' individual performances in two formats* (N = 12)

Student	One-on-one M	Paired M	Believed	Actual
1	7	7.5	One-one	Paired
2	10.5	9	One-one	One-one
3	9	11	Paired	Paired
4	10.5	12.5	Paired	Paired
5	7	7.5	One-one	Paired
6	11.5	10	One-one	One-one
7	10	8.5	One-one	One-one
8	9	10	Paired	Paired
9	10	7	One-one	One-one
10	10	10	Paired	No change
11	7	6.5	One-one	One-one
12	6.5	6.5	Paired	No change

Note. Score out of 16 points. M = Mean. The believed outcome was taken from students' PTQ part two.

Chapter V: Discussion

Research Question 1: Student Perceptions of Test Format

The first research question asked about student perceptions of the one-on-one oral test with the teacher and the paired oral test with another student; the goal being to gain insight into the test-takers' experience. The results were analyzed in terms of seven perceptual themes: *nervousness, preparedness, interest, interaction, effectiveness of format, belief in performance,* and *preference.* It was discovered that, generally, the two test formats yielded the nearly the same student outlook and feelings. However, there were slight differences in mean scores from the post-test questionnaires that correlated with comments made in the post-test questionnaire, part 2 commentary that reveal small patterns or insight worth discussing.

Nervousness in the one-on-one test format. A pattern worth noting was that students felt more nervousness before and during the one-on-one test than the paired test. The results of the post-test questionnaires and the post-test questionnaire, part 2 show an increase in nervousness when testing with the teacher. The comments made in the post-test questionnaire, part 2 commentary confirmed this as well. Three students made comments about being nervous in the one-on-one test and more comfortable in the paired test format. This pattern indicates that something about the one-on-one format caused students to have more nervousness.

A possible reason for the increased nervousness, as Bachman and Palmer (2010) suggest, may be because students felt threatened by the interaction with the teacher, caused by the imbalance of power in the relationship. Kormos (1999) refers to this kind of relationship in a conversational setting to be an "unequal social encounter" in which the interviewer has dominance in the conversation (p. 164). Van Lier (1989) describes the possibility of the one-on-one test as being more like an "interrogation", which would likely

cause stress to the student (p. 496). Interestingly, student eight commented about the social encounter in the paired test format by saying, “In this test I felt more comfortable, I think because I was talking with person he is like my level in English.” This comment suggests that by speaking with an equal rather than an authoritative figure may relieve nerves in an oral test.

Another cause of nervousness could be that because students generally perceived the one-on-one format to be more effective in representing their English ability, so they took it more seriously than the paired test format. In terms of students’ perceptions concerning the effectiveness of the formats, the data results revealed that students believed the one-on-one format to be slightly more effective in evaluating their English ability. Student two commented of the one-on-one format, “I think it is better to show the students’ real level of English skill…” Student six also commented, “I think it’s best way to know students’ English level.” Perhaps because students felt the test format to be more authentic, they were more concerned about their performance. It would seem that they care more about showing their ability to the teacher than to their peer.

The results showed that having nerves before or during the one-on-one test did not necessarily indicate poor performance in that format. Studies done by Park and Lee (2005), and Phillips (1992) measured the correlation between nervousness and performance and found that nerves had a negative effect on performance outcome. The only evidence of nerves negatively affecting performance was in the case of students one, three, and eight. They each indicated at some point as being more nervous in the one-on-one format. In the post-test questionnaire, part 2, student one commented on being nervous in both formats. Student three commented that in the one-on-one format, “…I was really nervous.” And in the paired format, “…I felt comfortable.” Student eight commented that in the one-on-one format, “I felt nervous a little bit.” And in the paired format, “I felt more comfortable.” The outcome

was that all three of these students performed worse in the one-on-one test than in the paired test. It appears that in these three cases, nervousness negatively affected their performance.

However, there were three other students who reported being more anxious in the one-on-one test, and yet they performed better in that format. The other two students who reported being more anxious in the one-on-one test received the same score in both formats. Overall, these results do not confirm that nervousness results in poor test performance.

Preparedness. The results did not reveal a significant pattern or difference in the students' view of test preparedness between the two formats. There was a slight (0.50) increase in feelings of preparedness in the paired format, but overall, students indicated a perception of being prepared in both formats. It can be speculated that one of the reasons for this was because students had been previously exposed to the two test environments earlier in the semester. In addition, throughout the course of the semester, students frequently participated in discussing topics with the teacher individually and with peers. The topics chosen for the prompts were taken from themes discussed in class. Previous experience with the test environment and the prompt topics may have contributed to the students' overall feeling of preparedness in both test formats.

There was one student who commented on lack of preparedness in both formats due to the prompt topic. Student three said, "I didn't choose the topic I know well" in the one-on-one test, and "I was not prepared for this topic" in the paired format. It may be speculated that this student had not personally prepared for the topics in the tests but studied other topics that were discussed in class instead. Another possibility is that the student may have felt that the discussions in class were not enough to prepare them for an oral test. This student's reaction, however, did not appear to correlate with other student's perceptions of preparedness.

Another student commented on the difficulty in preparing for a test with another student, as you cannot predict what they will say in advance. Student ten said of the paired test, "…it's hard to prepare for it because you don't know what others said." This comment leads to the possibility that the test with the teacher was somehow more predictable, and therefore easier to prepare for. It brings up the issue of the two test environments as being different conversational interactions or discourse structures. Van Lier (1989) describes the basic characteristics of conversation to be, "…unplannedness (local assembly), unpredictability of sequence and outcome, potentially equal distribution of rights and duties in talk, and manifestation of features of reactive and mutual contingency" (p. 495). Student ten's comment about the lack of predictability indicates that the paired format represents a more authentic conversation in which the two students must respond to one another spontaneously and equally, rather than an interview where the teacher is eliciting responses.

Interaction in the two test formats. Another interesting aspect of interaction reported on was about accommodation in the one-on-one format. As outlined previously, accommodation occurs when the interviewer (teacher in this case) facilitates the interaction. Ross (2007) refers to this as "scaffolding" the interaction, and attributes it to inconsistency in scoring and low inter-rater reliability (p. 2017). In the case of this study, the results show that although accommodation does not affect student scores, it does have an impact on student perceptions of the interaction. Student four commented, "I think it is a nice test because when I cannot talk or my brain stopped, the teacher asked me [a question], then I can talk." Student twelve commented, "When I forget what I should say, the teacher will help me." These comments suggest that during the one-on-one interaction, the teacher made the conversation easier for the students by helping them speak when they could not. Brown (2003) found that interviewers in the one-on-one format are intimately "implicated in the construction of

candidate proficiency" (p. 1). By accommodating the students, the teacher may have influenced the students' performance outcome in the one-on-one test format.

In contrast, in the paired test, the students reported on some of the difficulties of the interaction. Student five commented, "I think talk with another student have a little difficult because we can't describe clearly what we want to say, so we pause during the conversation." Student six said, "Sometimes, I don't understand partner and another student's speech. I can't say 'I don't understand', so just listen to continue. That's the problem." These comments reveal that students find it more difficult to cope in a conversation with their peer when there is a lack of understanding or an information gap. This follows Brooks' (2010) findings that, "in the paired test, the interaction was much more complex and revealed the co-construction of a more linguistically demanding performance" (p. 341). In a paired test setting, the students must rely more heavily on their own or their partner's interactional competence to fill in the pauses or ask questions to keep the conversation going, whereas in the one-on-one test format the teacher is relied on to facilitate the interaction.

As mentioned by Brooks (2009) previously, oral interaction is a performance "co-constructed among the participants" (p. 342). If building an oral performance is a joint experience, it seems, as evidence from student comments, that the teacher is being relied on by students to do more of the conversation building in the one-on-one test format. And, although students appear to find the paired interaction more difficult, it is more of an accurate reflection of classroom practices where peer discussion is more common than individual discussions with the teacher. The differences in the interactional nature of the one-on-one test versus the paired test are noteworthy and have implications for ESL classroom instruction and assessment.

Interest in the paired test format. Previous studies (Fulcher, 1996; Shohamy et al, 1986; Van Moere, 2006) have shown that students view paired oral tests more positively than

one-on-one tests. In contrast, the results in this study were that students generally reported being interested in both of the formats, with a slight increase in interest in the one-on-one format. Although the results found in this study do not directly confirm the previous studies, they still add some valuable insights to the conversation.

An interesting pattern that emerged while analyzing the data was that other feelings may have affected or been associated with students' overall concept of interest in this study. Students may have had negative or positive experiences in one format that caused them to be disinterested in that format or the other format. For example, four students commented on being nervous in the one-on-one test, but more comfortable in the paired test. In the end, they all reported being more interested in the paired test format. Being less nervous in the paired test format may have contributed to those student's being more interested in that format. Perhaps they were able to engage in the interaction or express themselves more due to being less nervous. Two other students reported on the difficulties of the interaction in the paired format, and then reported more interest in the one-on-one test format. They may have felt more interested in the one-on-one format because they could more easily express themselves. It is possible that nervousness and difficulty in the interaction may have factored into those students' perceptions of interest in the test formats.

One takeaway is that the perceptions that students have about the test formats may be difficult to isolate and measure independently. Students may also have a different concept of what "interest" means to them altogether and this needs to be accounted for. This requires more advanced design of questionnaire items that are better at specifically eliciting target perceptions. To gain a more comprehensive perspective of each perception, in-depth interviews with students may be necessary as well.

Self-belief in performance. In the background survey, approximately 67% of students who indicated belief in performing better in one of the tests, did perform better in

that test, and the other approximately 33% had no change in scores across formats. This reveals that students' belief in their ability to be successful in a specific performance, or self-efficacy beforehand, may have contributed to their actual success in the performance. In a study on the effects of self-efficacy, Schunk and Swartz (1993) confirm the relevance of the connection by reporting that, "…self-efficacy is positively related to skillful performance." (p. 11). Previous experience in both formats may have built confidence in those students who felt they would perform better in a specific format. Perhaps those students (~33%) who reported not knowing which test they would be more successful in did not have especially high self-efficacy in either format.

After the two tests were completed, approximately 67% of students correctly reported which test they achieved a better score in. After reflecting on their performance in each of the tests, it appears that a majority of students knew which format yielded their best performance. This indicates that a majority of students sense the differences in interactions and the components that make an interaction successful and/or unsuccessful. In-depth post-test interviews would be necessary for a deeper look into what students know about successful interactions.

Preference of format. Students indicated an exact divide in preference between formats, and two students preferred both test formats equally. A noticeable pattern was that there was a direct correlation between students' preferred test format and the test format they felt they received a higher score in. In the post-test questionnaire, part 2, all ten (100%) of the students who indicated a preference for a specific format, also reported a belief that they received a higher score in their preferred test format. It is unsurprising that students prefer the test format in which they feel they can be successful. In a test setting, every learner desires being set up in an environment where there is likely a rewarding outcome.

Another noticeable pattern was that four students (~33%) indicated a positive relationship between preference, belief in higher score, interest, and preparedness while anxiety was shown to be inversely related to their preference of format. This means that contributors to student preference of format include other factors like interest in the format, preparedness for the format, anxiety in the format, and how well they feel they can perform.

Research Question 2: Student Performance Scores

The second research question asked about student performance outcome in the one-on-one oral test versus the paired oral test, with the goal being to discover whether or not different test formats would affect the outcome of student performance. The results showed that there was only a slight increase in performance outcome in the paired test versus the one-on-one test, but it was not statistically significant. Therefore, it can be concluded that generally there was no difference in performance outcome between the formats. These results do not confirm the findings of Brooks (2009), in which participants in her study performed better in the paired test than the one-on-one test.

A possible explanation for the difference in outcome between the two studies has to do with familiarity of surroundings. The participants in this study have had exposure to the testing environments and the people they were testing with previously in the semester. Students tested with me, the classroom teacher, and their classmates whom they have known throughout the semester. We have had countless occasions to interact in discussion and become familiar with each other. In contrast, the participants in Brooks' (2009) study were testing with an unspecified examiner in the one-on-one interview and then an unspecified other participant in the paired test. The unfamiliarity may have contributed to participant test anxiety and other qualities which, in turn may have influenced performances. Taking differences of familiarity into account, the difference between the two studies' score outcomes may not be a meaningful or valuable comparison.

Pedagogical Implications

This study demonstrates that students have generally positive perceptions of both the one-on-one and paired test formats, and that they generally score the same in both formats. In light of these results, it seems that teachers should confidently use one or both test formats in their classroom assessment practices. However, teachers must also keep in mind that, even though the test formats yielded similar outcomes in scores and perceptions, they are not equal tests in terms of interaction.

From the results of student commentary, we can conclude that the one-on-one format and the paired format are two different tests. The presence of teacher accommodation in the one-on-one test format is evidence of a difference in interaction that is not present in the paired test format. The difficulty expressed by students in the paired format is evidence that there was a different range of interactional features the students had to rely on to communicate with a peer, versus with the teacher. Teachers should not consider the two test formats to be interchangeable, as they are not equal evaluations of student proficiency.

Considering these are two different tests, teachers should utilize them for different purposes in their assessment practices. Since the teacher plays the dominant role in the one-on-one test, it would be a useful format for eliciting specific information from students for evaluation like target vocabulary or grammar, or elaboration of ideas. In this case, the teacher has the power to steer the conversation towards a goal if necessary. If teachers do incorporate this format into their assessments, they should bear in mind that it is more nerve-racking for students than the paired test. Teachers should be intentional about finding ways to communicate with students individually throughout the semester and not just at testing times. This may eventually ease nerves for students in the one-on-one format.

Although there are some advantages to using the one-on-one format, teachers should primarily use the paired test format, as it reflects actual classroom and real-world conditions

(Bonk and Ockey, 2003; Brooks, 2009; Brown, 2003; Kormos, 1999; Ross, 2007; van Lier, 1989; Van Moere, 2006). In an ESL course where students are frequently interacting with their peers in communication, an oral assessment should be implemented to accurately evaluate the interactional skills that they use like turn-taking, negotiation of meaning, prompting elaboration, asking questions, managing a topic, etc. (Brooks, 2009). Webb (1994) suggests that pair or group testing is the fairer way of assessing students who are immersed in a collaborative classroom environment interacting with their peers.

Chapter VI: Conclusion

Limitations

One of the investigation's aims was to find out student perceptions of the two testing formats and to gain that insight through the information obtained in the background survey and post-test questionnaires. The environment of the two testing formats was designed to recreate a real classroom test setting. Throughout the testing sessions and afterwards, unforeseen flawed areas of the methodology became clear. It is possible that those design flaws may have affected the overall results of the data collection. In this section, I will discuss those limitations discovered and their possible effects on my research.

Role of assessment setting. The goal when designing the testing environment was to simulate real tests that the students in that particular course had taken before. The major difference in the tests designed for this research and the tests for the course was that these research tests were not for a grade, i.e. the outcome of the tests would not affect the students' course grade. This may have influenced students' attitudes before, during, and after the tests, and, in turn, affected the results of the scores and the questionnaires in some way. My attention was drawn to this possibility when interacting with the students during the data collection.

During the one-on-one data collection process, two different students, upon seeing their speaking prompt, directly expressed dislike for the topic and asked to change prompts. They both communicated that they did not know what to discuss regarding the topic. Had this been a test for their course grade, I speculate that they would not have reacted outwardly with such bold opinions.

When filling out the first post-test questionnaire, another student expressed an opinion concerning the items regarding nervousness. He told me that he was not nervous because this was "not a real test". He did not know how to answer the item on the questionnaire

accurately. This situation, and the above situation regarding the attitude towards the prompts, made me aware of the difficulty in designing a test setting that feels real for the students participating. When taking a test in which performance outcome influences course grade, there is a natural nervousness that students feel. When that nervousness is absent, how is their performance altered? How closely and accurately can I relate the data that I collected from the simulated tests to real, classroom-based assessment?

Role of background survey. The goal when designing the background survey was to obtain personal information like gender, native language, and topics of interests. The background survey was also used as an instrument to gather preliminary perceptions about the two test formats in regard to nervousness and belief in performance. During the course of the data analysis, I realized that too much unnecessary information was gathered from the background survey. The focus of the background survey should have been to gather more preliminary perceptions of the two test formats, rather than obtain personal information.

Effects of questionnaires. The goal when designing the questionnaires was to extract as much perceptual information from the students as possible immediately after they had completed each test. One limiting aspect of that goal was that the questionnaire may have been too lengthy and filled with redundant items. There were fifteen items in the post-test questionnaire and then seven items in the questionnaire provided after both tests had been completed. I found that some of the items were eliciting the same information. For example, item number seven in the post-test questionnaire states, “There were many interruptions from the other person,” item number eight states, “I could speak when I wanted to speak,” and item number eleven states, “I wanted to say more, but I missed my opportunity because other people were talking.” The wording is all different in these items, but they are basically drawing the same information from the student. As revealed in the results section, having different items eliciting the same information proved problematic because student answers

for the same thematic item had contradictory scores. These items could have been combined in a way to provide more efficiency in elicitation.

Another limitation to the questionnaires was the lack of order and intentionality in the topic groupings. The items were not grouped in a logical way to represent an easily extractable theme like *nervousness* or *belief in performance*. Instead, the item themes were randomized in a way to make it difficult to group them reasonably. Also, some of the groupings made when analyzing the results seemed to be underrepresented, like *interest*, which has only one item. I also found that there lacked any items regarding the prompt and/or topic of the prompt. Item number three in the background survey elicited what interest students had in topics to discuss, and yet there lacked any follow-up items in the post-test questionnaire eliciting feelings about the prompt topic.

Role of raters and inter-rater reliability. Two graduate assistants in the TESOL program were recruited to participate in scoring the oral assessments. Both raters had varying amounts of experience in instructing which may have affected the outcome of scores. One rater was a novice ESL teacher while the other was seasoned in the field. Although the scores may have been influenced by their conceptual understanding of the scoring scale and oral assessment in general, it does not appear that the scores were obviously different from rater to rater.

The two raters of the assessments did not complete inter-rater reliability measures. Before the raters scored the oral tests, I briefed them on the format procedure and environment. Then I reviewed the scoring scale with each of them, instructing them to strictly follow the descriptions in the scoring scale. Although all of this was done, the raters did not have the opportunity to calibrate their scoring by practicing with sample tests. Although there were not major discrepancies between each raters' scores, the lack of practice may have affected the results of the test scores, making them less reliable.

Recommendations for Future Research

To advance this study and to do a closer examination of student perceptions, more in-depth and specific feedback from students is necessary. This would help clarify and confirm the results of this study. One way this could be done is to improve the pre- and post-test questionnaires. Mentioned as a limitation, the questionnaires did not include enough items, or the items were too vague or underrepresented for students to properly express their perceptions of the test formats. Another way to examine perceptions more closely would be to do in-depth interviews with the students after completing one or both of the tests. This would allow students to elaborate on their impressions of each test format.

Another approach to future research may be to look more specifically at the effect of the *interlocuter*, the person with whom one is interacting with in the assessment, and whether or not that person's personal characteristics have an effect on the other's performance. In this study, the teacher was a female and, since the participants self-selected their pairs, there were mixed-gendered as well as same-gendered pairs. The pairs were also comprised of students from various language backgrounds and different ages. As mentioned in previous sections, these varying personal characteristics may play a role in altering their partner's performance, and while this study did not investigate those effects, it would be valuable for future research.

Another avenue of investigation is on the effect of the topic of the prompt in relation to test perceptions and performance. Since neither the pre-test or post-test questionnaires included items which elicited perceptions of the prompt topic, it is unclear whether or not the prompt topic played a role. The prompt topic in this study was about fear and the various effects it has on us and society in general. Could speaking about fear induce students to be more fearful or nervous? What are the unintended consequences of the prompt topic and how do they influence student test performance?

A final suggestion for future research may be to look more closely at paired tests. As an assessment that is potentially more fair and reflective of real world and classroom conditions, the paired test format warrants further study. Future research may be done to investigate the correlation between interactional features taking place within the classroom versus the interactional features taking place within the paired test format and one-on-one test format. It would be valuable for classroom teachers to know how the interactional features compare in each setting. This would encourage teachers in the classroom to intentionally point out specific features of interaction that need more attention or provide specific feedback to students trying to improve upon real-world communication. By discovering which test more accurately reflects what happens in the classroom, teachers can improve their oral assessments and potentially prevent harmful test backwash among students.

References

Bachman, L. & Palmer, A. (2010). *Language assessment in practice.* Oxford, England: Oxford University Press.

Berwick, R. & Ross, S. (1992). The discourse of accommodation in oral proficiency interviews. *Studies in Second Language Acquisition, 14(2),* 159-176.

Bonk, W.J., & Ockey, G.J. (2003). A many-faceted Rasch analysis of the second language group oral discussion task. *Language Testing, 20(1),* 89-110.

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-constructing a better performance. *Language Testing, 26,* 314-366.

Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing, 20(1),* 1-25.

Cheng, L., Rogers, T., Hu, H. (2004). ESL/EFL instructors' classroom assessment practices: Purposes, methods, and procedures. *Language Testing, 21(3),* 360-389.

Creswell, J., Klassen, A., Clark, V., & Smith, K. (2010). *Best practices for mixed methods research in the health sciences.* National Institutes of Health. Retrieved November 5, 2017, from https://www2.jabsom.hawaii.edu/native/docs/tsudocs/Best_Practices_for_Mixed_Methods_Research_Aug2011.pdf

Davis, L. (2009). The influence of interlocutor proficiency in a paired oral assessment. *Language Testing, 26(3),* 367-396.

Frost, K., Elder, C., & Wigglesworth, G. (2011). Investigating the validity of an integrated listening-speaking task: A discourse-based analysis of test-takers' oral performances. *Language Testing, 29(3),* 345-369.

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

Hughes, A. (2003). *Testing for language teachers*. Cambridge, England: Cambridge University Press.

Jones, E.E., & Gerard, H.B. (1967). *Foundations of social psychology.* New York: John Wiley & Sons, Inc.

Kormos, J. (1999). Simulating conversations in oral-proficiency assessment: A conversation analysis of role plays and non-scripted interviews in language exams. *Language Testing, 16(2),* 163-188.

Leaper, D.A., Riazi, M. (2013). The influence of prompt on group oral tests. *Language Testing, 3,* 177-204.

Lumley, T., & O'Sullivan, B. (2005). The effect of test-taker gender, audience and topic on task performance in tape-mediated assessment of speaking. *Language Testing, 22(4),* 415-437.

McNamara, T.F. (1997). 'Interaction' in second language performance assessment: Whose performance? *Applied Linguistics, 18(4),* 446-466.

Nitta, R., & Nakatsuhara, F. (2014). A multifaceted approach to investigating pre-task planning effects on paired oral test performance. *Language Testing, 31(2),* 147-175.

Ockey, G.J. (2009). The effects of group members' personalities on a test taker's L2 group oral discussion test scores. *Language Testing, 26(2),* 161-186.

O'Loughlin, K. (2002). The impact of gender in oral proficiency testing. *Language Testing 19(2),* 169-192.

O'Sullivan, B. (2000). Exploring gender and oral proficiency interview performance. *System, 28,* 373-386.

Phillips, E. M. (1992). The effect of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal, 76(1),* 14-26.

Park, H., & Lee, A. R. (2005). L2 learners' anxiety, self-confidence and oral performance. Paper presented at The Pan-Pacific Association of Applied Linguistics (PAAL), Japan.

Ross, S.J. (2007). A comparative task-in-interaction analysis of OPI backsliding. *Journal of Pragmatics, 39,* 2017-2044.

Schunk, D. H. & Swartz, C.W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, Georgia.

Shohamy, E., Reves, T., & Bejarano, Y. (1986). Introducing a new comprehensive test of oral proficiency. *English Language Teaching Journal, 40,* 212-220.

Song, J. (2014). *A study of ESL students' performance and perceptions in face-to-face and virtual-world group oral tests* (Dissertation). Retrieved November 5, 2017, from The University of Texas at Austin Library Repository.

Taylor, L., & Wigglesworth, G. (2009). Are two heads better than one? Pair work in L2 assessment contexts. *Language Testing, 26(3),* 325-339.

van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral proficiency interviews as conversation. *TESOL Quarterly, 23,* 489-508.

Van Moere, A. (2006). Validity evidence in a university group oral test. *Language Testing, 23(4),* 411-440.

Webb, N. M. (1995). Group collaboration in assessment: Multiple objectives, processes, outcomes. *Educational Evaluation and Policy Analysis, 14(2),* 239-261.

Young, R., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition,* 403-424.

Appendix A: Pre-Test Background Survey

Background Survey

Name:

Please read each statement and mark the option that applies to you. The survey takes about 5 minutes to complete. Thank you for your time.

1. **Gender:** Male _____ Female _____
2. **What is your native language?**
3. **What kinds of topics do you like to talk about in discussion activities? (circle the letters)**
   a. Science and technology
   b. Traditions in different cultures
   c. Popular culture & entertainment
   d. Society & politics
   e. Personal experiences & jobs
   f. Controversial issues
   g. Travels & adventures
   h. Hobbies & personal interests
   i. Other: _____ _____
4. **What kinds of activities do you like? (circle the letters)**
   a. Individual work
   b. Pair work/discussions
   c. Small group work/discussions
   d. Individual presentations
   e. Pair/group presentations
   f. Interviewing people from outside of class
   g. Research & presentation
   h. Read/listen/watch & debate
   i. Free conversation
   j. Games
   k. Other: _____ _____
5. **Which test makes you more nervous? (circle one letter)**
   a. Speaking test with the teacher
   b. Speaking test with another student
6. **Which test do you think you can do better on? (circle one letter)**
   a. Speaking test with the teacher
   b. Speaking test with another student

Appendix B: Speaking Prompts

What Fear Can Teach Us Prompt

*You will have a discussion with your teacher or another student. You will read the text and discuss the topic. You are expected to discuss for **5 minutes.***

Talk about a time you were scared. Did you learn anything from that fear? What positive or negative effects do our fears have on us? What other emotions, such as happiness or anger, can also teach us something? Support your view.

.

You will begin the discussion in one minute.

Fear and Media Prompt

*You will have a discussion with your teacher or another student. You will read the text and discuss the topic. You are expected to discuss for **5 minutes.***

Describe some common fears that society has. What types of stories that focus on fears do you often see in the media? In your experience, do media stories often make situations seem worse than they really are? Do you think the media spreads fear in people? Support your answers.

You will begin the discussion in one minute.

Appendix C: Post-Test Questionnaires

Oral Test Questionnaire – After 1st and 2nd Test

Name:

*Please read each statement and put an **X** the option that applies to you. The questionnaire takes about 15-20 minutes to complete. Thank you for your time.*

1 **10**

(Strongly Disagree) ---------- (Strongly Agree)

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. I felt nervous before the test.										
2. I felt ready for this test style (with a teacher or another student).										
3. I was nervous during the test.										
4. I felt comfortable speaking with the other person (teacher or another student).										
5. It was interesting to do the test in this environment (with the teacher or with a partner).										
6. It was easy to take turns during the test.										
7. There were many interruptions from the other person.										
8. I could speak when I wanted to speak.										
9. I could explain my ideas well.										
10. I could ask questions easily.										
11. I wanted to say more, but I missed my opportunity because other people were talking.										
12. I found it easy to say what I wanted to say during the conversation.										
13. This was a good test to show my English skills.										
14. I showed my real level of English conversation ability in this test.										
15. I believe I did well on the test.										

Appendix D: Post-Test Questionnaire, Part 2

Oral Test Questionnaire Part 2 – After 2nd Test

*Please read each question and put an **X** in the box you think is best.*

	Test with Teacher	Test with Another Student
1. In which test do you think you got a higher score?		
2. Which test made you feel more nervous?		
3. Which test was more interesting?		
4. Which test do you think you were more prepared for?		
5. Which test style do you prefer?		

1. What is your opinion about the one-on-one test with your teacher? Please explain.

2. What is your opinion about the test with another student? Please explain.

Appendix E: Scoring Scale

Scoring Scale

Score	Delivery	Language Use	Topic Development	Interactional Competence
0	Fragments of speech that are so halting that conversation is not really possible. Sounds incomprehensible.	Cannot produce a sentence.	Topic is not developed at all.	Shows no awareness of other speakers; may speak but not in a conversation-like way.
1	Consistent pronunciation and intonation problems cause considerable listener effort and frequently obscure meaning. Delivery is choppy, fragmented. Speech contains frequent pauses and hesitations.	Produces very basic sentence forms. Overall, turns are short, structures are repetitive, and errors are frequent.	Limited relevant content is expressed. The response lacks substance beyond expression of very basic ideas. Speaker may be unable to sustain speech to complete the task.	Does not initiate interaction, produces monologue only; Shows some turn-taking, may say, “I agree with you,” but not relate ideas in explanation; too nervous to interact effectively.
2	Speech is clear at times though it exhibits problems with pronunciation, intonation or pacing and so may require significant listener effort.	Primarily uses basic sentences; more complex-structures are absent or contain significant errors. Vocabulary sufficient to discuss topic, but generally simple. Errors are common.	The response is connected to the task; though the number of ideas presented, or the development of ideas is limited. Mostly basic ideas are expressed with limited elaboration.	Response to others without long pauses to maintain interaction; shows agreement or disagreement between others’ opinions.

3	Speech is generally clear with some fluidity of expression, but it exhibits minor difficulties with pronunciation, intonation or pacing and may require some listener effort. Overall intelligibility remains good.	Produces a mix of short and complex sentence forms, typically uses shorter forms. Vocabulary is adequate to discuss topics at length. Errors in grammar and vocabulary are sometimes noticeable.	Response is coherent and sustained and conveys relevant ideas. Overall development is somewhat limited.	Generally confident responds appropriately to others' opinions. Shows ability to negotiate meaning quickly and naturally.
4	Speech is clear, fluid and sustained. It may include minor difficulties with pronunciation. Pace may vary at times. Overall intelligibility remains high.	Makes use of longer sentences and a variety of structures. Uses a range of vocabulary; words are precise. Errors remain but not distracting.	Response is sustained and sufficient to the task. It is generally well developed and coherent; relationships between ideas are clear.	Turn-taking is very smooth. Can initiate discussion and conclude the discussion. Shows agreement and disagreement with the interlocutors.

Appendix F: Background Survey Results

Results of topics of interest

Item	Topics	Result
Q3 What kinds of activities do you like to talk about in discussion activities? (circle answers)	Science & technology	0
	Traditions in different cultures	3
	Popular culture & entertainment	4
	Society & politics	1
	Personal experiences & jobs	3
	Controversial issues	0
	Travels & adventures	3
	Hobbies & personal interests	8
	Other: Art	1
	Other: Nutrition & sports	1

Note. N = 12. Students instructed to circle as many topics as they wanted.

Results of activities of interest

Item	Activities	Result
Q4 What kinds of activities do you like? (circle answers)	Individual work	1
	Pair work/discussions	1
	Small group work/discussions	5
	Individual presentations	2
	Pair/group presentations	0
	Interviewing people	1
	Research & presentation	1
	Read/listen/watch debate	4
	Free conversations	5
	Games	6

Note. N = 12. Students instructed to circle as many topics as they wanted.

Item 5 and 6 results (N = 12)

Item	One-on-one	Paired	Misc. answers
Q5 Which test makes you more nervous?	6	2	Neither – 4
Q6 Which test do you think you can do better on?	4	2	I don't know – 4 No difference – 2

Appendix G: Post-Test Questionnaires Results

Descriptive statistics of the one-on-one and paired post-test questionnaire results (N = 12)

Item	Mean (SD)	
	One-on-one	Paired
Q1 I felt nervous before the test	4.17 (2.89)	3.75 (2.61)
Q2 I felt ready for this test style	6.92 (1.68)	7.42 (1.93)
Q3 I was nervous during the test	4.67 (3.73)	3.50 (2.67)
Q4 I felt comfortable speaking with the other person	7.50 (2.43)	8.08 (1.88)
Q5 It was interesting to do the test in this environment	8.42 (2.27)	9.08 (1.00)
Q6 It was easy to take turns during the test	7.50 (2.24)	7.83 (2.04)
Q7 There were many interruptions from the other person	4.25 (3.65)	3.83 (3.56)
Q8 I could speak when I wanted to speak	7.33 (2.46)	7.58 (2.50)
Q9 I could explain my ideas well	7.00 (2.17)	7.17 (2.79)
Q10 I could ask questions easily	6.58 (2.61)	7.42 (1.62)
Q11 I wanted to say more, but I missed my opportunity	2.83 (3.10)	3.50 (3.48)
Q12 I found it easy to say what I wanted to say	7.67 (1.97)	7.25 (2.53)
Q13 This was a good test to show my English skills	9.17 (1.11)	8.42 (1.83)
Q14 I showed my real level of English	7.50 (1.98)	7.25 (1.96)
Q15 I believe I did well on the test	6.58 (2.47)	7.08 (2.47)

1: Strongly disagree – 10: Strongly agree

Appendix H: Post-Test Questionnaire, Part 2 Results

Post-test questionnaire, part 2 results (N = 12)

Item	One-on-one	Paired	Misc. answers
Q1 In which test do you think you got a higher score?	7	5	
Q2 Which test made you more nervous?	4	2	Both – 1, Neither – 5
Q3 Which test was more interesting?	6	4	Both – 2
Q4 Which test do you were more prepared for?	6	5	Neither – 1
Q5 Which test style do you prefer?	5	5	Both – 2

Appendix I: Post-Test Questionnaire, Part 2 Commentary Results

Student	One-on-One	Paired
1	*I'm nervous, but comfortable because I can adjust my emotions when teacher asked me.*	*Also nervous, but it is interesting. We can share our ideas for each other.*
2	*I think it is better to show the students' real level of English skill, and it can let teacher know their shortcoming, then help them improve their abilities.*	*I think it is very interesting, that I never try to do this way.*
3	*I think it was a good practice, but I was really nervous. I didn't choose the topic I know well.*	*I think it was not preferred, but we give some interesting informations. I was not prepared for this topic, but I felt comfortable when I answered.*
4	*I think it is nice test because when I cannot talk or my brain stopped, the teacher asked me, then I can talk.*	*It is nice idea to talk with student.*
5	*I think is very comfortable when I talk to my teacher. I don't feel nervous. I think it was a pleasant conversation.*	*I think talk with another student have a little difficult because we can't describe clearly what we want to say, so we pause during the conversation.*
6	*I think it's best way to know students' English level. Also, it can help students to learn English successfully. Because, during the conversation, I learned English from teacher (like how she says question, sentence).*	*Sometimes, I don't understand partner and another student's speech. I can't say "I don't understand", just listen to continue. That's the problem.*
7	*I like it because it helps you improve your English skills and builds up your confidence.*	*It helped me in knowing other ideas from someone else and helped me improving my English.*
8	*It was a good test, but I felt nervous a little bit. I don't know why.*	*In this test I felt more comfortable, I think because I was talking with person he is like my level in English.*
9	*No opinion.*	*For test with another student, we always have same point. This not interesting. I prefer we have difference opinion debate that.*
10	*I feel super nervous about one-on-one test. I always forget the things I prepare before. However, it is interesting to have a conversation with teacher. It can help me explain my idea well.*	*I feel relax, just like a normal conversation. I don't feel so nervous. However, it's hard to prepare for it because you don't know what others said. But interesting.*

Appendix I Continued *Post-test questionnaire part 2 commentary results*

Student	One-on-One	Paired
11	*I think you are really good, so I don't have any opinion.*	*Asking more questions about the text.*
12	*When I forget what I should say, the teacher will help me.*	*Just ask the question and feel relax.*

Note. N = 12

Appendix J: IRB Approval

ST. CLOUD STATE UNIVERSITY
EDUCATION FOR LIFE

Institutional Review Board (IRB)
720 4th Avenue South AS 210, St. Cloud, MN 56301-4498

Name: Hayley Miller

Email: hamiller@stcloudstate.edu

IRB PROTOCOL DETERMINATION: Expedited Review-1

Project Title: Students Perceptions and Performances in Two Oral Assessment Formats

Advisor Shawn Jarvis

The Institutional Review Board has reviewed your protocol to conduct research involving human subjects. Your project has been: **APPROVED**

Please note the following important information concerning IRB projects:

- The principal investigator assumes the responsibilities for the protection of participants in this project. Any adverse events must be reported to the IRB as soon as possible (ex. research related injuries, harmful outcomes, significant withdrawal of subject population, etc.).

- For expedited or full board review, the principal investigator must submit a Continuing Review/Final Report form in advance of the expiration date indicated on this letter to report conclusion of the research or request an extension.

-Exempt review only requires the submission of a Continuing Review/Final Report form in advance of the expiration date indicated in this letter if an extension of time is needed.

- Approved consent forms display the official IRB stamp which documents approval and expiration dates. If a renewal is requested and approved, new consent forms will be officially stamped and reflect the new approval and expiration dates.

- The principal investigator must seek approval for any changes to the study (ex. research design, consent process, survey/interview instruments, funding source, etc.). The IRB reserves the right to review the research at any time.

If we can be of further assistance, feel free to contact the IRB at 320-308-4932 or email ResearchNow@stcloudstate.edu and please reference the SCSU IRB number when corresponding.

IRB Chair:

Dr. Benjamin Witts
Associate Professor- Applied Behavior Analysis
Department of Community Psychology,Counseling, and Family Therapy

IRB Institutional Official:

Dr. Latha Ramakrishnan
Interim Associate Provost for Research
Dean of Graduate Studies

OFFICE USE ONLY

SCSU IRB# 1761 - 2231	**Type: Expedited Review-1**	**Today's Date: 2/5/2018**
1st Year Approval Date: 2/5/2018	**2nd Year Approval Date:**	**3rd Year Approval Date:**
1st Year Expiration Date: 2/4/2019	**2nd Year Expiration Date:**	**3rd Year Expiration Date:**