

Preliminary Study of the Role of Eye Contact, Gestures, and Smiles Produced by Chinese-as-a-first-language Test-takers on Ratings Assigned by English-as-a-first-language Examiners during IELTS Speaking Tests

by

Christiani Pinheiro Thompson

B.A, B.Ed., Universidade Federal Do Rio de Janeiro, 1991

A Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of

MASTER OF ARTS

in the Department of Linguistics

© Christiani Pinheiro Thompson, 2016
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy or other means, without the permission of the author.

Supervisory Committee

Preliminary Study of the Role of Eye Contact, Gestures, and Smiles Produced by Chinese-as-a-first-language Test-takers on Ratings Assigned by English-as-a-first-language Examiners during IELTS Speaking Tests

by

Christiani Pinheiro Thompson

B.A., B.Ed., Universidade Federal do Rio de Janeiro, 1991

Supervisory Committee

Dr. Li-Shih Huang, Department of Linguistics
Supervisor

Dr. Martha McGinnis-Archibald, Department of Linguistics
Departmental Member

Abstract

Supervisory Committee

Dr. Li-Shih Huang

Supervisor

Dr. Martha McGinnis-Archibald

Departmental Member

This study investigated the role of gestures, smiles, and eye contact on scores assigned to English-as-an-additional-language (EAL) speakers during standardized face-to-face speaking tests. Four English-as-a-first-language examiners and four EAL test-takers participated in simulated IELTS Speaking Tests. Qualitatively, an inductive thematic analysis was conducted. Quantitatively, scores were holistically (overall scores assigned) and analytically (by criterion). Nonverbal cues were examined by the total number of cues produced by all test-takers, the frequency of production by test-taker, the frequency of production of subcategories of nonverbal cues by test-taker, and by production alongside speech or in isolation. Mimicry of nonverbal cues generated by test-takers was investigated. Test-takers' lexical range was also analyzed vis-à-vis the scores assigned to the criterion *lexical resource*. Conclusions drawn from the triangulation of data sources indicate that nonverbal cues may have played a role in the assessment of the criteria *fluency and coherence* and *pronunciation*. This study adds to the current body of literature on second language assessment, which has suggested that variables other than language proficiency may play a role in scores assigned to test-takers during face-to-face speaking tests.

Table of Contents

| | |
|--|------|
| Supervisory Committee | ii |
| Abstract | iii |
| Table of Contents | iv |
| List of Tables | vi |
| List of Figures | vii |
| Acknowledgments | viii |
| Dedication | x |
| Chapter 1: Introduction | 1 |
| Chapter 2: Literature Review | 4 |
| 2.1 Intercultural Communication | 4 |
| 2.1.1 Ethnocentrism | 5 |
| 2.1.2 Zero-Acquaintance Situations | 7 |
| 2.2 Nonverbal Communication | 8 |
| 2.2.1 Eye Contact/Face Gazing | 10 |
| 2.2.2 Gestures | 12 |
| 2.2.3 Smiles | 15 |
| 2.3 Communicative Competence and Language Proficiency Assessment Tests | 18 |
| 2.4 Language Proficiency Tests | 20 |
| 2.5 Interview-Testing Format and Video-Stimulated Verbal Recall | 21 |
| Chapter 3: Methods | 25 |
| 3.1 Research Design | 25 |
| 3.2 Participants | 26 |
| 3.3 Instruments | 27 |
| 3.3.1 Examiners' Background Questionnaire | 28 |
| 3.3.2 Test-takers' Background Questionnaire | 30 |
| 3.3.3 Modified Rating Sheet | 31 |
| 3.3.4 The IELTS Speaking Test | 32 |
| 3.4 Video-stimulated Verbal Recall Sessions | 34 |
| 3.5 Pilot Study | 36 |
| 3.6 Main Study | 37 |
| 3.6.1 Recruitment | 37 |
| 3.6.2 Data Collection Procedures | 38 |
| 3.7 Transcriptions | 41 |
| 3.7.1 Transcription of verbal data | 41 |
| 3.8 Coding | 41 |
| 3.8.1 Nonverbal cues | 42 |
| 3.8.2 Video stimulated verbal recall sessions | 49 |
| 3.9 Speaking Tests | 53 |
| 3.10 Field Notes | 56 |
| 3.11 Data Analysis | 57 |
| Chapter 4: Results | 61 |
| 4.1. Test Scores | 61 |
| 4.2. Lexical Analysis | 67 |
| 4.3 Thematic Analysis and Frequency of Nonverbal Data | 69 |

| | |
|---|-----|
| 4.3.1 Perception of test-takers' overall behaviour. | 69 |
| 4.3.2 Frequency of production of test-takers' nonverbal cues..... | 71 |
| 4.3.3 Perception of test-takers' eye contact. | 73 |
| 4.3.4 Frequency of production of test-takers' eye contact..... | 75 |
| 4.3.5 Perception of test-takers' gestures. | 76 |
| 4.3.6 Frequency of production of test-takers' gestures..... | 78 |
| 4.3.7 Perception of test-takers' smiles. | 82 |
| 4.3.8 Frequency of production of test-takers' smiles..... | 84 |
| 4.3.9 Encouragement through nonverbal cues..... | 86 |
| Chapter 5: Discussion..... | 90 |
| 5.1 Key Findings..... | 90 |
| 5.1.1 Observable nonverbal cues vis-à-vis test-takers' performances..... | 90 |
| 5.1.2 Scores per criterion..... | 96 |
| 5.1.3 Examiners' perceptions of nonverbal cues vis-à-vis test-takers' performances..... | 98 |
| 5.2 Implications..... | 100 |
| 5.2.1. Methodological Implications..... | 100 |
| 5.2.2 Pedagogical and Practical Implications..... | 101 |
| 5.3 Limitations..... | 102 |
| 5.3.1. Study Design..... | 102 |
| 5.3.2. Setting..... | 103 |
| 5.3.3 IELTS Speaking Test and Examiners..... | 103 |
| 5.3.4. Video-stimulated Verbal Recall Sessions..... | 103 |
| 5.3.5 Data Analysis..... | 104 |
| 5.3.6 Nonverbal Cues..... | 105 |
| 5.4 Future Research Directions..... | 105 |
| 5.5. Conclusion..... | 106 |
| References..... | 108 |
| Appendices..... | 127 |

List of Tables

| | | |
|----------|--|----|
| Table 1 | <i>Examiners' Characteristics</i> | 28 |
| Table 2 | <i>Countries in Which Participants Worked as EAL Instructors and as Examiners</i> | 29 |
| Table 3 | <i>Examiners' Questionnaire Answers - Thoughts about Chinese-as-a-first- language Test-takers' Performances</i> | 29 |
| Table 4 | <i>Test-takers' Characteristics</i> | 31 |
| Table 5 | <i>Description - IELTS Speaking Test</i> | 33 |
| Table 6 | <i>Data Collection Sessions</i> | 40 |
| Table 7 | <i>Sample of an Ongoing Coding Session of Nonverbal Data (Examiner F and Test-taker 6)</i> | 44 |
| Table 8 | <i>Video-stimulated Verbal Recall Sessions - Themes, Codes, Examples, and Observations</i> | 52 |
| Table 9 | <i>Face-to-face and Audio-recorded Performances (per Test-taker)</i> | 62 |
| Table 10 | <i>Audio Recorded Tests - Mean Scores and Standard Deviation by Criterion</i> .. | 63 |
| Table 11 | <i>Mean Scores Assigned by all Examiners to Audio-recorded Tests – Scores by Criterion</i> | 64 |
| Table 12 | <i>Median Scores by Criterion</i> | 66 |
| Table 13 | <i>K-words - Families, Types, and Tokens</i> | 68 |
| Table 14 | <i>Frequency of Observable Nonverbal Cues per Test-taker</i> | 72 |
| Table 15 | <i>Frequency of Observable Nonverbal Cues per Test-taker (in Proportion to the Total Cues Produced per Test-taker)</i> | 72 |
| Table 16 | <i>Frequency of Observable Mutual Eye Contact in Isolation/alongside Speech</i> 76 | |
| Table 17 | <i>Frequency of Overall Observable Gestures by Type</i> | 79 |
| Table 18 | <i>Frequency of Observable Gestures by Type per Test-taker</i> | 80 |
| Table 19 | <i>Frequency of Gestures in Isolation (I/I)</i> | 81 |
| Table 20 | <i>Frequency of Gestures alongside Speech (A/S)</i> | 81 |
| Table 21 | <i>Frequency of Production of Smiles by Type</i> | 84 |
| Table 22 | <i>Frequency of Observable Duchenne and non-Duchenne Produced in Isolation (I/I)</i> | 85 |
| Table 23 | <i>Frequency of Observable Duchenne and Non-Duchenne Smiles Produced alongside Speech (A/S)</i> | 86 |
| Table 24 | <i>Nonverbal Cues Mimicked by Examiners</i> | 89 |

List of Figures

| | |
|---|----|
| <i>Figure 1.</i> Design of the study. | 26 |
| <i>Figure 2.</i> Example of object-adaptor. | 47 |
| <i>Figure 3.</i> Example of self-adaptor. | 47 |
| <i>Figure 4.</i> Example of illustrator..... | 47 |
| <i>Figure 5.</i> An example of the organization of the themes. | 50 |

Acknowledgments

Many wonderful people have supported me throughout my degree. First and foremost, I would like to thank Dr. Li-Shih Huang for all the support, encouragement, and guidance provided in my undergraduate and graduate years. Her belief in this study, her countless feedbacks, and her unquestionable knowledge made this degree possible. I would also like to thank Dr. Martha McGinnis-Archibald for her insight, kindness, and support, and to my external committee member, Dr. Tsung-Cheng Li, for his interest in the study and for the contributions made to this thesis.

There are several other people who should also be acknowledged for the amazing support they have offered me. First of all, I want to thank my family for all the patience they have had (you are amazing, thank you!). I also want to thank Jenny, who was always ready to offer support and help in moments of need. I would like to acknowledge the positive influence all the faculty members in the department have had in my educational career (including those with whom I did not have the opportunity to study but who would take the time to talk to me and listen to my ideas). In special, I want to thank the professors with whom I have worked, Dr. McKercher, Dr. Kirkham, and Dr. D'Arcy. I will be forever grateful for the guidance and the invaluable opportunities you have given me to learn and grow.

I have met many wonderful people both in and through the department, and some of them have become an important part of my life. Thank you Sky, Marianne, Janet, Catherine, Carrie, Amjad, Diane, Raj, Jesse, Lisa, Paul, Tess, Yiran, and Emily for

always being there for me. You guys are amazing! Cynthia Korpan and Moussa Magassa, I will never be able to thank you enough for your support, insight, and kindness.

There are many others who are not part of the university who have done a lot to help me complete my degree. Anne Gilbody, thank you for suggesting I should go back to school, for supporting me all the way through, and for being such a great boss; Pam Rubidge, I don't even know how to thank you for all you have done; Robynne, Michelle, Fred, Patricia, Allie, Gyneth, Karen, and Laurie-Anne, thank you for being there for me and for caring so much, and Nick Bartzis - thank you for always believing in me and for never doubting I could do it.

Dedication

I dedicate this thesis to my sons, Anthony and Spencer, to my siblings, Cláudia and Luciano, and to my parents, Linda and Elias, who have not only instilled the passion for learning in all their children but have also cheered, supported, and celebrated each one of our accomplishments. Vocês são maravilhosos.

Chapter 1: Introduction

What you do speaks so loudly I cannot hear what you say. (Emerson, 1883)

The International English Language Testing System (IELTS), a language proficiency test taken by over 2.7 million English-as-an-additional-language (EAL) learners yearly in more than 140 countries (IELTS, 2016), is a gatekeeping standardized test used to assess test-takers' chances of achieving a successful performance in higher education institutions. The IELTS Speaking Test is a one-to-one interaction between an examiner and a test-taker in which the examiner measures a test-taker's language proficiency during a face-to-face interview (IELTS, 2016). Review of the literature on speaking tests shows that exchanges between examiners and test-takers and scores assigned to test-takers may be affected by factors other than language proficiency (Briegel-Jones, 2014; O'Loughlin, 2000; O'Sullivan & Lu, 2002). Studies have also shown that factors such as gender and accent may affect exchanges between examiners and test-takers and the rapport established between them (Briegel-Jones, 2014; Merrylees, 1999; O'Loughlin, 2000; O'Sullivan & Lu, 2002); however, the possible role of nonverbal cues on scores assigned by examiners remains underexplored.

A review of the literature on second language assessment reveals that limited attention has been given to the role of nonverbal communication on face-to-face exchanges among individuals from different cultural backgrounds (Macan, 2009). This argument is supported by the common absence of identification of participants' cultural backgrounds in research (e.g., Krumhuber et al., 2009; Liden, Martin, & Parsons, 1993; Washburn & Hakel, 1973). Despite the importance of oral exchanges in intercultural communication and the awareness of the many layers that comprise such interactions, the primary focus of most studies in the field of intercultural communication has been on the lexicon and the syntactical structures of English as produced by EAL speakers

during exchanges with English-as-a-first-language (EL1) speakers (He, 2012; Hirvela, Nussbaum, & Pierson, 2012; Hudelson, 1989; McKay, 1994; Read & Nation, 2002).

From Cicero's (Hall, 2004) and Quintilian's (Kendon, 1997) first known studies on hand use in oratory, research has examined different aspects of nonverbal behaviour, such as Darwin's (1872) scientific analysis and consequent argument of the universality of expression of emotions or the investigation of individuals and the environment in which nonverbal cues are produced (Ambady, Hallahan, & Rosenthal, 1995; Burgoon & Poire, 1999; Imada, & Hakei, 1977). Studies on nonverbal features such as gestures, posture, eye contact, pauses, tone, and facial expressions have shown that intercultural interviews still present a challenge since nonverbal features are commonly misunderstood and may negatively affect assessments of candidates during employment interviews (Barrik, Shaffer, & DeGrassi, 2009; Campbell & LeVine, 1961; Neuliep, Hintz, & McCroskey, 2005). Research has also shown that factors such as ethnocentrism may influence the perception of nonverbal behavior (Samovar, Porter, McDaniel, & Roy, 2013; Swami & Furnham, 2012).

The noticeable dearth of studies that focus on how EAL speakers' nonverbal communication may affect ratings assigned during speaking tests conducted by EL1 examiners supports new research that may provide valuable information for examiners, test-takers, and those responsible for the design of language proficiency tests. Findings from such studies are relevant to higher education institutions worldwide, including Canadian universities, which have seen the number of international students more than triple since 1995 (AUCC, 2011). In a culturally diverse world, awareness of culture-specific nonverbal features may provide information that could be valuable for the assessment of language proficiency test outcomes.

Through an in-depth analysis, my intention was to develop an initial understanding of the role of three nonverbal cues (gestures, smiles, and eye contact), which have been identified by the

literature as having an influence on interviewers' perception of interviewees (Hall, 2004; Krumhuber & Manstead, 2009; McGovern & Tinsley, 1978; Schilbach et al., 2006) on scores assigned during standardized speaking tests. Empirically, the results of this study contribute to the understanding of the role of gestures, smiles, and eye contact in the context of language assessment and may also provide IELTS examiners and EAL test-takers with valuable information on the role these features may play in intercultural exchanges during the IELTS speaking test. Practically, the results can inform the development of a training program that can raise examiners' and test-takers' awareness of nonverbal features.

Chapter 2: Literature Review

To support this study, an overview of elements that are fundamental to studies on interpersonal communication are presented. First, intercultural communication and its relevance to the field of communication research are examined followed by a discussion of two factors that have been found to impact intercultural communication and perception of nonverbal communication in intercultural exchanges (Gifford, 1994; House, 2001; Swami & Furnham, 2012). Next, communicative competence and language proficiency tests are discussed. The final section will discuss the use of interview testing format and video-stimulated verbal recall as procedures for collection of nonverbal data.

2.1 Intercultural Communication

Language has been described as “the most explicit type of communicative behavior that we know of” (Sapir, 1935, p. 105). Although communication is inherent to interactions among all individuals and intercultural exchanges have become everyday occurrences worldwide with the aid of globalization, technology (e.g., the internet, airplanes), and through immigration, verbal and nonverbal channels have only been systematically researched in intercultural communication since the 1960s (Samovar & Porter, 2003).

Exchanges among individuals of distinct cultural backgrounds have been the topic of numerous studies (Durant & Shepherd, 2009; Liddicoat, 2009; Washington, Okoro, & Thomas, 2012), which have mainly focused on the analysis of the influence of culture on the perception of messages being exchanged. As argued by Samovar and Porter (2003), the importance of studies that focus on intercultural communication is supported by the understanding that “culture helps govern and define the conditions and circumstances under which various messages may or may not be sent, noticed, or interpreted” (p. 7).

In a multicultural world, understanding the cultural and sociocultural rules that are expressed through individuals' communicative acts is crucial. This is especially relevant when generalizations may result from misinterpretations during such interactions. The complexity of the relationship between communication and culture requires individuals to be aware of possible biases, overgeneralizations, and stereotyping (Vrij, 1993). Durant and Shepherd (2009) argued about the dangers of stereotypes playing a role in interactions among individuals of different cultural backgrounds, especially due to the changeable nature of such exchanges. In intercultural communication, awareness that the receiver may not possess the cultural and social values of the sender is primordial as communication relies heavily on the perception and the perspective of the individual who is receiving the message. Several factors, such as ethnocentrism and zero-acquaintance settings, may influence the perception of messages being exchanged (Barrick, Swider, & Stewart, 2010; Neuliep, Hintz, & McCroskey, 2005). These factors can lead to communication breakdown, misunderstanding, and even to inaccurate perception of deception.

2.1.1 Ethnocentrism

Ethnocentrism, the belief that an individual's cultural group is superior to other groups, is one of the core elements of studies on intercultural communication. LeVine and Campbell (1972) stated that, although ethnocentrism was initially perceived as a sociological concept, it was eventually seen as a psychological construct, "a widespread tendency of individuals to identify with their own in-group and simultaneously reject the out-groups" (Pocovnicu & Vasilache, 2012, p. 478). Perceiving one's culture as centrally important was succinctly defined by Hofstede, who argued that "ethnocentrism is to a people what egocentrism is to an individual" (as cited in Neuliep, Hintz, & McCroskey, 2005, p. 43). Being an essential element to culture preservation, it has been presented as a variable in studies on intercultural communication and

nonverbal perception in intercultural exchanges (Samovar, Porter, McDaniel, & Roy, 2013). The influence of ethnocentrism is not only limited to the perception of both verbal and nonverbal cues (Neuliep & McCroskey, 2005) - it also affects how observers see the source producing such cues. Knowledge of ethnocentric views in intercultural exchanges is essential as it acknowledges individuals' unique perceptions of their culture when compared to others'. Ethnocentrism may regulate the perception of nonverbal cues produced by individuals from different cultural and sociocultural backgrounds, which could result in biased perceptions.

Concerns about cultural bias in interview settings led to a call for studies on ethnocentrism in cross-cultural exchanges in interview settings (Campbell & LeVine, 1961). The need for investigation of the role of individuals' cultural backgrounds in interview settings was also supported by House (2001), who argued that similarities between interviewers and interviewees could affect hiring decisions as interviewers are more likely to hire candidates with whom they share cultural and/or ethnic values. The possibility that ethnocentrism may negatively impact the assessment of test-takers' language proficiency and, consequently, the scores that are assigned to them, must be explored as a lack of awareness of possible cultural biases may result in unfair assessments. If out-groups are seen as inferior to ethnocentrics when compared to in-groups and if ethnocentrism affects the perception of attractiveness and credibility (Neuliep, Hintz, & McCroskey, 2005), it is possible that biased rating decisions may take place during oral assessment tests. Such questions underlie and support further investigation of features that could affect the reliability of ratings assigned in settings that involve individuals who do not share the same ethnicity or cultural backgrounds.

2.1.2 Zero-Acquaintance Situations

Influence of nonverbal cues in zero-acquaintance situations, those in which unacquainted individuals judge each other's personality in the absence of social interaction (Albright, Kennedy, & Malloy, 1988), has also been discussed in nonverbal communication studies (Ambady, Hallahan, & Rosenthal, 1995; Kenny, Horner, Kashy, & Chu, 1992; Swami & Furnham, 2012). A study by Albright, Kennedy, and Malloy (1988) presented findings that stress the strong influence of nonverbal cues in judgments of personality attributes. This argument was supported by Ambady, Hallahan, and Rosenthal (1995), whose study focused on inferences made by individuals with minimal to no prior contact. Their findings showed that social skills and extraversion can be accurately assessed even when individuals do not know each other.

In the context of interviews, initial perceptions of individuals can affect interviewers' decisions and assigned ratings. One example is Barrick, Shaffer, and DeGrassi's (2009) study, which showed that handshakes and smiles can directly influence impressions made by candidates. Barrick, Swider, and Stewart (2010) found similar results in their study, further supporting the argument that initial impressions not only affect candidates' ratings but may also predict final hiring decisions.

Within the testing context, previous studies on factors that can hinder the perception of nonverbal cues produced by EAL speakers during exchanges with EL1 speakers suggest that a deeper understanding of their role is still necessary. Furthermore, findings suggest that misunderstanding of nonverbal cues produced by EAL speakers may result in unfair assessments. Since lack of knowledge of cultural differences in the use and/or perception of nonverbal cues may be reflected in scores given to test-takers, in-depth studies may provide a better insight into their role in environments in which individuals may be judged in absence of previous interaction.

2.2 Nonverbal Communication

In 1872, Charles Darwin's *The Expression of the Emotions in Man and Animals* presented the first scientific study on nonverbal behaviour. Although his work has been deemed valuable for presenting the first functional analysis of nonverbal behaviour, it failed to acknowledge that nonverbal communication may be influenced by culture (Gudykunst & Ting-Toomey, 1988). Since then, several studies have explored the perception of nonverbal cues (Burgoon, Buller, & Woodall, 1989; Ekman & Friesen, 1969; Ekman & Rosenberg, 1997; Hall, 1978; Hall, 1984; Rosip & Hall, 2004), with most focusing on the study of the individual and/or the interaction, that is, the environment in which nonverbal cues are produced (Ambady, Hallahan, & Rosenthal, 1995; Imada, & Hakel, 1977; Burgoon & Poire, 1999).

Nonverbal communication, a multi-channel process that is constituted of elements other than verbal language, can be classified as static (such as appearance, including skin colour and attire), dynamic (such as gestures, smiles, eye contact), and paralinguistic ("the features that accompany speech and contribute to the meanings people assign to the overall transaction") (Samovar, Porter, McDaniel, & Roy, 2013, p. 291). Nonverbal cues convey messages that can be intentional or unintentional, and due to their irrepressible nature and connection to emotion representation, have been researched in areas such as anthropology, psychology, education, and business. They reinforce, substitute, or contradict verbal behaviour and transmit information about individual's personalities. As Gifford (1994) stated, individuals' personalities are encoded in nonverbal behaviour. Samovar et al. (2013) support Gifford's statement by explaining that nonverbal cues "serve to express internal states, create identity, regulate interaction, and substitute for words" (p. 291).

Nonverbal behaviour may be encoded differently from culture to culture (Gifford, 2006).

Nonverbal cues may express different messages to different cultures, making nonverbal behaviour an important element of studies in intercultural communication. Perception and production of such elements rely heavily on an individual's culture; the continuum of nonverbal behaviour, many times produced in conjunction with verbal messages, clarify meaning in the context in which they are produced. As a result, knowledge of the cultural subtleties and meanings associated with nonverbal cues requires those taking part in intercultural communicative exchanges to be aware of the influence of each other's cultures on their communicative acts. As Singelis (1994) stated:

The fact that at least one communicator is working in a second language means the verbal content may not be as clear as it would be in an intra-cultural interaction. Consequently, the reliance on nonverbal communication may be even greater than normal. (p. 275)

In intercultural exchanges, nonverbal behaviour may be even more relevant when language difficulties arise. The importance of raising awareness about the influence of nonverbal cues in intercultural exchanges results from the knowledge that potential miscommunication in exchanges between individuals from different cultures is high. Nonverbal cues such as gestures, eye contact/face gazing, and smiles have been shown to affect both the perception of individuals and the message they try to convey, and their relevance can be confirmed by studies throughout history (Kendon, 1997) that have focused on the role of nonverbal cues in communicative acts.

2.2.1 Eye Contact/Face Gazing

The eyes are commonly referred to as “the window to the soul” and, in interactions, are used to determine others’ emotional and/or mental state (Adams, Nelson, & Purring, 2013, p. 229). The eyes and the region that surrounds them have been shown to express genuine expressions of emotions more than any other part of the face (Buck, 1988). Research has shown that eye contact is used to show a desire for approval (Efran, 1968) and to demonstrate sincerity (Exline et al., 1970), among other functions. Samovar, Porter, McDaniel, and Roy (2013) explained the relevance of eye contact/face gazing to studies of communication by stating that “eyes express emotions, monitor feedback, indicate degrees of attentiveness and interest, regulate the flow of the conversation, influence changes in attitude, define power and status relationships, and help modify impression management” (p. 286). Direct eye contact has been identified as a sign of self-confidence by interviewers (Tessler & Sushelsky, 1978) and also as a demonstration of honesty and sincerity (Exline et al., 1970). A study by Tessler and Sushelsky (1978) focused on the implications of eye contact behaviour in employment interview settings, and results showed that participants who did not make eye contact were perceived as less confident.

In studies of eye contact (pupil to pupil), the use of an eye-tracking device has been considered a useful tool for proper collection and analysis of data as the eyes “provide an ideal and powerful objective measure of ongoing cognitive processes and information requirements during behaviour” (Tatler et al., 2014, p. 3); however, the discomfort caused by its use can also negatively impact the interaction between interviewees and interviewers and, possibly, the overall production of nonverbal cues (Harrigan, 2013). Although eye-to-eye contact is considered an important aspect of interpersonal interactions (Tessler & Sushelsky, 1978), research has shown that head direction may have a stronger impact than direct eye-to-eye contact (Duncan & Fiske,

1977; Exline & Fehr, 1982), suggesting that studies on face gazing (looking towards an individual's face) rather than on eye gazing (pupil to pupil) may provide researchers with valid data. Exline and Fehr (1982) have addressed questions concerning the reliability of data collection of gaze by arguing that live interactions and video-taped records usually provide high reliability estimates. This argument was supported by Schilbach and his colleagues' (2006) study on gaze direction and perception of engagement between participants and a virtual agent. In their study, participants who were directly gazed at identified more engagement with the virtual agent than those who interacted with virtual agents that looked at another person. Similarly, despite arguments that have questioned the validity of measurement of gaze contact, studies by Argyle (1970) and Kendon (1970) have shown that measurement is possible as participants tend to either look away or look at each other's faces during interactions, suggesting that precise data on eye-to-eye contact may be less critical (Harrigan, 2013).

Studies such as Tessler and Sushelsky's (1978) on the effect of eye contact on perceptions of job applicants do present valuable findings in the role of eye contact/gaze in exchanges; however, the lack of information on the cultural background of participants may present a limitation. For instance, a study by Woo and Prud'homme (1999) examined the role of eye contact in Chinese negotiations; findings suggested that, during interactions with Chinese individuals, eye contact should be avoided as it caused discomfort and could be perceived as a sign of disrespect. Direct eye contact is also considered a sign of disrespect in some Latin cultures (Samovar, Porter, McDaniel, & Roy, 2013), and avoidance of direct eye contact may be misinterpreted by individuals from other cultural backgrounds who are unaware of the influence of culture on the production of this nonverbal gesture.

Failure in acknowledging cultural influences on the production and perception of direct

gaze may affect the validity of studies on eye-contact/face gazing¹ due to the many variables that are inherent to intercultural exchanges that may not be taken into consideration during data analysis.

2.2.2 Gestures

Gestures are “primarily hand movements [. . .] that are used basically for two purposes – to illustrate speech and convey verbal meaning” (Matsumoto & Hwang, 2013, p. 75). Sapir (1949) compared gestures to a secret code that, despite being understood by the members of a cultural group, are not guided by any explicit rules (Rosenthal, 1978). Although not all gestures are understood by all individuals, the universality of nonverbal cues as an elaborate coding system shows that they are as much part of communication as language (Knapp & Daly, 2002).

Nonverbal cues are an inherent part of the communicative act, and gestures of different types are present in most exchanges; they convey messages and communicate through a visual medium (Bull & Doody, 2013). According to Sapir (1951), “one may intuitively interpret the relatively unconscious symbolisms of gesture as psychologically more significant in a given context than the words actually used” (p. 79).

Gestures have received attention since ancient times, with the first known studies being those by Quintilian in AD 100 (Kendon, 1997) and Cicero (Hall, 2004), who analyzed the role of hand use in oratory. Since then, research, such as Darwin’s (1872) study on the possible universality of shoulder shrugs, has focused on both production and perception of gestures. By the 20th century, researchers concentrated on different roles of gestures as they tried to uncover connections such as those between gestures and language origin (Wells, 1987) and gestures and

¹ Note that although this research investigates “face gazing,” the term “eye contact” will be used henceforth.

speech (Bull & Doody, 2013; Cassell, 1998). Gestures are an essential component of communication, and their importance is highlighted by the fact that, even though verbal language can be filled with speech errors and hesitations, gestural errors are quite uncommon; moreover, during exchanges, speakers rely on the information being shared through gestures to complement the verbal message and, at times, to even correct speaker's errors (McNeill, 1992).

Gestures can be classified as *emblematic*, the symbolic gestures that translate a message known to most members of a subculture (such as head nods) (Ekman & Friesen, 1977), *illustrators*, gestures that are connected to or accompanied by speech that illustrate, strengthen the verbal message, and are usually produced unintentionally (such as pointing to an object or depicting an action) (Harrigan, 2013), *adaptors*, those produced with little awareness that are used to cope with feelings, thus having an affective function (such as scratching the head or rubbing hands) (Ekman, 1977), and *regulators*, which are used to maintain the flow of conversation (such as head nods and postural shifts) (Ekman, 2004).

Since Efron's studies on culture and the production of gestures (Efron, 1941), researchers have investigated the influence of culture on the perception of gestures in intercultural exchanges (Ekman, 1976; Friesen, Ekman, & Wallbott, 1979) and have argued that some may be classified as culture-specific. A shoulder shrug, for example, is seen as a sign of confidence in the Middle East but means the opposite in most Western countries (Jokinen & Allwood, 2010), while a thumbs-up, which is understood as a representation of 'good' in Western countries, is considered obscene in the Middle East (Knapp & Hall, 2006). Gestures have also been linked to individuals being perceived as more composed, more competent (Maricchiolo et al., 2009) and have been found to influence observers' preference as individuals who produce gestures are usually liked more than those who do not (Kelly & Goldsmith, 2004).

Data on gestures can be collected via three main approaches: the data glove based

approach, which consists of participants wearing a glove-like device with sensors that identify movements of fingers and hand; the vision-based approach, which makes use of video cameras to register hand movements and is considered a natural way to collect data; and the colour glove based approach, a somewhat combination of the two first approaches, in which participants wear coloured gloves to make visual data collection simpler (Zhu, Yang, & Yuan, 2013). Video recording has been identified as a useful tool to collect reliable nonverbal data as it provides researchers with the opportunity to review interactions and gestures produced, resulting in a higher level of accuracy (Harrigan, 2013). Questions, however, have been raised regarding the reliability of gesture coding. Coding and reliability of data on gestures was discussed by Harrigan (2013), who proposed that researchers should make use of trained observers who are aware of well-determined parameters for identification of gestures and should also record behaviours and frequency of production in relation to the time they occur. This would then allow production of gestures to be linked to other parts of the interaction, such as greetings (Harrigan, 2013).

Despite their occurrence alongside language, review of the literature shows that gestures have received little attention in the field of bilingualism (Kendon, 2004), a factor that could result from cultural variation in how they are produced and perceived. Concerns also have been raised in regards to data collection (Gullberg, 2010) as irregularity in production and inconsistency in reaction times vary not only from individual to individual but also from context to context. To address these concerns, researchers have suggested that qualitative research principals can provide valuable information regarding the contexts (e.g., social, cultural, discursive) in which gestures may occur as well as an opportunity for researchers to observe natural production of gestures and perform an in-depth analysis of the data (Gullberg, 2010).

Culture undoubtedly permeates all communicative acts and the channels through which individuals communicate. Knowledge of the influence of individuals' cultural background on

gestures and how they are decoded by those who do not belong to the same sociocultural groups is essential for messages to be properly understood. Studies on gestures have shown that cross-linguistic analyses of gesture production and perception are still uncommon and support that qualitative studies based on observations of face-to-face interactions provide an opportunity for researchers to better understand the role of gestures in intercultural exchanges. In short, the lack of studies that focus on the role of gestures in intercultural exchanges in zero-acquaintance settings support further investigation that may provide valuable information on how EL1 individuals perceive gestures used by EAL speakers.

2.2.3 Smiles

Smiles are strong interpersonal cues that can be used to appease and apologize among other functions (DePaulo, 1992). Although smiles are universal acts, their production and perception are influenced by sociocultural rules (Samovar, Porter, McDaniel, & Roy, 2013). As Kraut and Johnston (1979) argued, culture influences smiling “both by determining the interpretation of events, which affects the cause of happiness, and by shaping display roles, which determine when it is socially appropriate to smile” (p.1540).

One of the smiles that has received attention in the past few years for being recognized as a positive display of emotion is the Duchenne smile, a facial expression that includes activation of a major muscle that pulls the lip corners up into a smile (the zygomatic muscle) and of the muscles that narrow the eyes and raise the cheeks to create crow’s feet around them (orbicularis oculi) (Ekman, Friesen, & Hager, 2002). The Duchenne smile, named after researcher Guillaume Duchenne (1862), is seen as a sign of extraversion and generosity (Mehu, Little, & Dunbar, 2007). It is also perceived as genuine especially when it is spontaneously produced (Krumhuber & Manstead, 2009). Duchenne smiles have also been shown to increase perception of

interviewees as more competent and hireable when compared to interviewees who produced non-Duchenne (also called false) smiles during interviews (Woodzicka, 2008).

The relevance of studies on the perception of facial expressions, such as smiles, was supported by Beattie, Webster, and Ross (2010), who argued that individuals spend more time looking at facial expressions displayed during interactions than looking at the gestures produced. Gunnery and Hall (2014) supported that argument and suggested that the perception of smiles and the resulting social outcomes need to be further investigated. Smiles are produced and decoded differently from culture to culture. Findings from Freisen's (1973) study, for example, showed that Japanese participants who initially displayed negative motions while watching stressful films smiled in the presence of an older male experimenter while American participants continued displaying their negative emotions in the same context. Cultural rules guide displays of emotions. Smiles can represent embarrassment in Japan (Nishiyama, 2000) and shallowness in Korea (Samovar, Porter, McDaniel, & Roy, 2013) but are commonly associated with happiness and considered a representation of interest and trustworthiness in North America (Samovar, Porter, McDaniel, & Roy, 2013).

Krys et al.'s (2014) study involving individuals from seven countries presented findings that mostly corroborate the positive perception of individuals who smile; however, they also argued that different cultures perceive smiles differently. In terms of perception of intelligence, for example, individuals who smile were seen as more intelligent in Hong Kong (Lau, 1982) but less intelligent in Japan (Matsumoto & Kudoh, 1993). Smiles also increase perception of self-esteem of those who produce them and may be seen as a signal of competence and self-confidence (Krys et al., 2014).

Measurements of spontaneously produced smiles, as with most displays of emotions, can be affected by the intentionality of the encoder in producing such cue. They may influence

observers' perception of emotions as the smile would not reflect the "true feelings" of the individual who is producing it (Buck & Power, 2013). Ekman and Friesen (1969) have suggested that unobtrusive observation and recording of interactions, especially during exchanges that do not require individuals to follow display rules, may minimize perception of deception and provide observers with more accurate information on the smiles being produced. Video recording, therefore, may provide researchers with a valuable opportunity to observe and identify smiles as well as other displays of emotion.

Research has also shown that smiles can be posed and may not be discriminated by observers (Hess & Kleck, 1994). Moreover, findings have shown that observation of posed or real smiles may not be related with perceived happiness of the individuals who produce them (Hess et al., 1989). Kappas, Krumhuber, and Küster (2013), however, have argued that such findings may be the result of stereotypes and biased perception. Although these studies present valuable concerns in regards to perception of smiles and their association with emotion, I believe that factors such as cultural background and gender could account for their findings. In zero-acquaintance settings, intercultural exchanges may be affected by ethnocentrism, which may result in erroneous perceptions of feelings and personality traits being transmitted unconsciously by individuals. The relevance of research on smiles and other nonverbal cues in zero-acquaintance settings acknowledges their communicative role in social interaction (see Yamamoto & Suzuki, 2006). As Murphy, Lehrfeld, and Isaacowitz (2010) stated, "accurately recognizing different smiles types would contribute to successful social interactions and effective communication" (p. 812). This knowledge validates in-depth studies that focus on how smiles are perceived in intercultural exchanges as they may provide relevant information that could potentially minimize negative effects of cultural biases in such environments.

2.3 Communicative Competence and Language Proficiency Assessment Tests

Language is an instrument used for social interaction and is governed by rules (Hymes, 1972). Historically, linguistic competence - the knowledge of grammatical form - was the focus of teaching and assessment instruments; however, the focus on grammatical form fails to address the proper evaluation of the knowledge involved in the use of language in real-life contexts; i.e., linguistic knowledge does not account for all the elements that constitute oral communicative competence. The understanding that verbal communication relies in interaction was discussed in Canale and Swain's (1980) influential theoretical framework of communicative competence, which consists of grammatical competence (such as the knowledge of grammar and syntax), sociolinguistic knowledge (related to the sociocultural rules of language use), discourse competence (related to the ability to produce cohesive and coherent spoken or written text in a language) and strategic competence (related to how speakers overcome issues when difficulties in communication arise). The ability to use language to interact with others was also discussed in Bachman and Palmer's (1996) model of language ability, which states that language ability consists of elements such as grammatical knowledge and knowledge of cultural references and figures of speech, which enable speakers to make the best use of their language knowledge (Bagarić, 2007).

Canale and Swain (1980) have argued that communicative competence and communicative performance are distinguishable as the latter refers to language use. This differentiation, in their opinion, should be addressed in language assessment tests by the use of tasks that elicit communicative performance from test-takers. Proficiency tests are constructed with the intention to evaluate examinees' language ability (Luoma, 2004). Ratings from proficiency tests should reflect test-takers' communicative competence, be fair, and be useful based on the purpose for

which they were constructed. For those to be achieved, developers focus on the reliability (the consistency of the scores) and the validity (the meaningfulness of the scores in regards to their intended use) of proficiency tests (Chalhoub-Deville & Turner, 2000; Young, So, & Ockey, 2013). To ensure that similar results would be assigned to the same test-takers if they were to take the same test more than once (reliability), instruments and procedures are of great importance.

The perception of nonverbal communication as part of communicative competence has been supported by researchers. For instance, Savignon (1972) argued that “communicative competence may be defined as the ability to function in a truly communicative setting - that is, in a dynamic exchange in which linguistic competence must adapt itself to the total informational input, both linguistic and paralinguistic, of one or more of the interlocutors” (p. 8). Young (2002) also argued that the role of nonverbal behaviour in oral proficiency tests has not yet been properly addressed (p. 255). As Huang (2013) stated, “much systematic research has examined the construct validation of the concept of communicative competence in L2 education” (p. 6); however, in the field of second-language testing, the role of nonverbal cues on ratings assigned to test-takers by examiners remains largely underexplored.

Most assessment tests evaluate test-takers’ communicative competence by measuring reading, writing, listening, and speaking skills. Of the four skills commonly assessed in proficiency tests, speaking is considered one of the most challenging domains as several factors may influence an examiner’s impression of a test-taker’s speaking ability (Luoma, 2004). This argument strongly supports in-depth analyses of features, such as nonverbal behaviour, that may affect ratings assigned during oral assessment tests.

2.4 Language Proficiency Tests

Reliability of results from language assessment tests such as the IELTS has been the focus of extensive research. Chalhoub-Deville and Turner (2000) argued that any variable other than language proficiency that can potentially affect tests scores could be reflected in errors of measurement and might result in limitations to the reliability and generalizability of scores assigned to test-takers. Chalhoub-Deville and Turner (2000) also stated the importance of empirical evidence and theoretical arguments that support the interpretation of scores, which should be an ongoing process, essential for the validity of the test to be determined. As a result, factors that might influence ratings assigned to test-takers have been examined (Briegel-Jones, 2014; Kang, 2012; Merrylees, 1999; O'Loughlin, 2000; O'Sullivan & Lu, 2002), with results indicating discrepancies in several aspects of the exchanges between examiners and test-takers that may affect scores given by the examiners. Studies have shown that examiners' styles are inconsistent (Kang, 2012; Merrylees, 1999; O'Loughlin, 2000; O'Sullivan & Lu, 2002). Findings also show differences in the rapport established with candidates (Briegel-Jones, 2014; Merrylees, 1999), examiners' support given to test-takers during speaking tests (O'Sullivan & Lu, 2002), and discrepancies in talking time between examiners and test-takers (Merrylees, 1999; O'Sullivan & Lu, 2002). Variables such as gender, accent, and examiners' exposure to foreign languages have also been investigated (O'Loughlin, 2000; O'Sullivan & Lu, 2002; Winke, Gass, & Myford, 2013), but the influence of nonverbal cues on exchanges during the IELTS Speaking Test remains mostly unexplored (Briegel-Jones, 2014).

As the number of international students rises in higher education institutions around the world, it is essential for researchers to focus on factors that may hinder examiners' proper assessment of test-takers' communicative competence as misjudgment directly impacts test-

takers' chances of pursuing university degrees at institutions that require students to be fluent speakers of English. Findings from research on second language assessment have indicated that variables other than communicative competence may influence the assessment of test-takers' performances. This argument supports in-depth investigations of nonverbal cues as one of the possible elements that could play a role in how examiners evaluate test-takers' performances during face-to-face language proficiency tests.

2.5 Interview-Testing Format and Video-Stimulated Verbal Recall

Since Darwin's (1872) use of static images to exemplify his description of emotions, pictures of individuals' facial displays of emotion have been used in research on nonverbal communication (Etcoff & Magee, 1992; Wilhelm et al., 2014). Despite findings showing that judgment of personality traits based on static images can be accurate (Naylor, 2007), some studies fail to take into account the context and other nonverbal cues that might have been concomitantly produced as they solely provide information of an isolated display of emotion captured at a moment in time (e.g., DeGroot & Gooty, 2009; Naylor, 2007). This limitation has resulted in interviews being considered a valuable procedure for collection of nonverbal data (DeGroot & Gooty, 2009; Gifford & Wilkinson, 1985; McGovern & Tinsley, 1978; Nguyen et al., 2014) as they offer researchers an opportunity to study real-time production of nonverbal cues in conjunction with oral production. The possibility of audio- and video-recording interviews may also be seen as another strength of this procedure as they provide researchers with full access to contextualized data.

In language proficiency assessment tests, the oral assessment component of the test is commonly a one-on-one interview between a test-taker and an examiner (Luoma, 2004). Despite some criticism in regards to the power examiners have over test-takers (Bachman, 1988), the

control that the examiner has over the interaction is considered one of the strengths of the interview format (Luoma, 2004). Research on face-to-face speaking tests is valuable to the field of communication because interviews are interpersonal interactions. Findings show that interviews are a good predictor of performance (Huffcutt, Roth, & McDaniel, 1996; Wiesner & Cronshaw, 1988). Based on the understanding that data collected from interviews can provide a deeper and broader scope of candidates' skills due to both verbal exchanges and displays of nonverbal behaviour, interviews have become a commonly used instrument for collection of data on nonverbal cues (Huffcutt, Roth, & McDaniel, 1996; Briegel-Jones, 2014).

The relevance of studies focusing on the role of nonverbal cues in interview settings is supported by the literature (Fiksdal, 1988; Huffcutt, Roth, & McDaniel, 1996; Lauer, 2005; Nguyen et al., 2014; Wiesner & Cronshaw, 1988), which shows a limited number of environments in which their role in intercultural communication has been studied. Moreover, many studies on both verbal and nonverbal communication in interview settings have not carefully examined the role of participants' cultural backgrounds in interview situations involving multicultural participants (Imada & Hakel, 1977; Liden, Martin, & Parsons, 1993; McGovern & Tinsley, 1978; Parton, Siltanen, Hosman, & Langenderfer, 2002). While some have identified that participants belonged to different cultures (Chen, Chen, & Lin, 2013; DeGroot & Gooty, 2009), studies do not always take into account participants' cultural backgrounds as a variable during analysis of data. New studies on such contexts, therefore, can clarify if culture-influenced perception of nonverbal cues affect ratings given by examiners during oral proficiency tests. Burgoon and Poire (1999) identified the influence of culture on perception of nonverbal cues and stated that "many nonverbal behaviours comprise a socially shared vocabulary analogous to verbal communication" that "would have recognizable interpretations by individuals of the same sociocultural group" (p. 107). That means that these behaviours may not be recognizable or may

be misinterpreted by those who belong to different sociocultural groups. Posthuma, Morgeson, and Campion (2002) suggested that interviewers may use nonverbal behaviours to make attributions about applicants, what stresses that awareness of cultural rules that may affect the use of nonverbal cues produced by test-takers and how they are perceived by examiners is fundamental for oral assessment tests to be properly evaluated.

As findings from previous studies suggest, intercultural exchanges during face-to-face assessments can provide valuable information on the role of nonverbal cues produced by EAL test-takers on ratings assigned by EL1 IELTS examiners (Burgoon & Poire, 1999; Chalhoub-Deville & Turner, 2000; Young, 2002). This study was an initial investigation on the role of three nonverbal cues that have been identified in the literature as having an influence on how individuals are perceived (gestures, eye contact, and smiles) produced by CL1 speakers on ratings during simulated IELTS speaking tests held by EL1 examiners, an area currently unexplored. The choice to use simulated IELTS speaking tests rests on the importance of findings being representative of real-life oral assessment processes. In an attempt to obtain an insight on the possible role of nonverbal cues on the scores assigned during the speaking tests, the scores assigned to the face-to-face and audio-recorded performances were compared as they could provide information on possible influence of nonverbal cues on examiners' perceptions and evaluation of test-takers' performances.

The relevance of studies on perception of nonverbal cues and the limited number of studies that have focused on exchanges between EAL and EL1 speakers in zero-acquaintance settings support the use of simulated speaking tests such as the one used in this study. Observation of production of nonverbal cues during face-to-face intercultural exchanges provides tools for an in-depth analysis of exchanges.

Having identified a gap, this study examined how EL1 IELTS examiners perceived and

rated EAL speakers in simulated IELTS speaking tests and presented a comprehensive analysis of the production of gestures, eye contact, and smiles (alongside speech or in isolation) generated during face-to-face tests.

This study intends to answer the following question:

- What is the role of gestures, eye contact, and smiles produced by Chinese-as-a-first language test-takers on ratings assigned by English-as-a-first-language examiners during simulated IELTS Speaking Tests?

Chapter 3: Methods

This chapter discusses the methods used in this study. First, the research design, participants, and instruments used are discussed followed by the description of data collection procedures, transcriptions, coding of the data, and data analysis.

3.1 Research Design

In this study, six CL1 graduate students registered at a university in Western Canada at the time of the study and EL1 examiners² participated in simulated IELTS Speaking Tests. The study consisted of two sessions. In the first session, six face-to-face simulated speaking tests, each consisting of one examiner and one test-taker, were conducted using a simulated speaking test (see Appendix A) retrieved from a website³ that offers preparatory courses to potential test-takers. The same simulated testing conditions and procedures were followed for all speaking tests. Immediately upon conclusion of each test, I conducted video-stimulated verbal recall sessions, which were audio and video recorded, with each examiner.

Two weeks after the conclusion of session one, session two commenced. In this session, all examiners evaluated the performances of all test-takers in the audio-recording of the face-to-face speaking tests. Figure 1 illustrates the overall design of the study (information on sessions is detailed in Section 3.7.1).

² The reasons for the choice of purposive sampling are detailed in Section 3.3.

³ Retrieved from <http://www.ipassielts.com>.

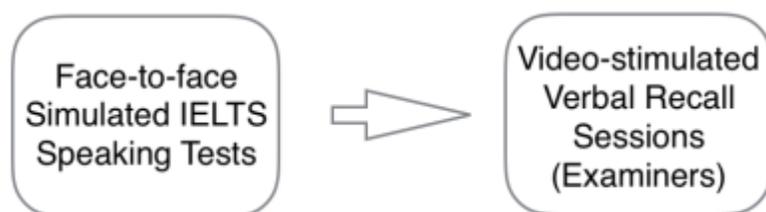
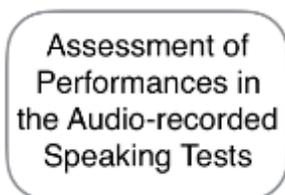
Session 1:**Session 2:**

Figure 1. Design of the study.

3.2 Participants

This study involved six EL1 examiners and six graduate students from the People's Republic of China as test-takers. The graduate students were, at the time of the study, pursuing a degree at a university in Canada.

The sample size and the selection of participants' cultural backgrounds reflect my decision to conduct an in-depth, multilayered analysis of the data. The criteria established for the selection of test-takers for this study were as follows:

1. Focusing on one cultural group of test-takers rather than on test-takers from different cultural backgrounds minimized culturally-influenced variables that might affect the production of nonverbal cues during the speaking test;
2. The decision to recruit CL1 speakers was based on the fact that they comprise one of the largest groups at the graduate level in most higher education institutions in Canada (Canadian Bureau for International Education - CBIE, 2015);

3. Recruiting graduate students limited the possibility of considerable gaps in test-takers' communicative abilities as all international graduate students admitted to the institution in which the study was conducted must receive a score equivalent to IELTS 6.0⁴.

In regards to examiners, changes in the contract between IELTS and certified examiners resulted in the original design of this study being modified. Instead of certified examiners, EL1 examiners with at least 10 years of experience assessing EAL students' speaking skills were invited to take part in the study. All examiners who took part in the study were familiar with the IELTS Speaking Test; examiners had either instructed IELTS preparation courses or had previously administered IELTS tests.

Of the twelve participants who took part in the speaking tests, four (examiner B/test-taker 2; examiner C/test-taker 3) were not included in the data analysis due to either procedural or technical problems. Two speaking tests, therefore, were omitted from the data analysis: examiner B/test-taker 2 did not follow the guidelines established to administer the speaking test, and examiner C/test-taker 3's video recording of the speaking test was corrupted.

3.3 Instruments

The following instruments were used to collect data for this study: examiners' background questionnaire, test-takers' background questionnaire, a sample IELTS Speaking Test, and a modified IELTS rating sheet.

⁴ The minimum score international graduate varies between 6.0 and 7.0 depending on the department which the students intend to join.

3.3.1 Examiners' Background Questionnaire

The examiners' background questionnaire (see Appendix D) was designed to gather information on their country of birth, number of hours a day spent with EAL learners, countries in which they have worked, experience as examiners, and their perception of CL1 test-takers. Examiners' experience ranged from six to over 12 years, with a mean of 11. They reported speaking English with EAL learners between three and six hours a day and having previously lived in countries such as Singapore, Japan, Hong Kong, and South Korea between less than one year and 25 years. Table 1 provides information about examiners' characteristics.

Table 1

Examiners' Characteristics

| | | |
|--|--------|---------|
| Age (in years) | Mean | 47.0 |
| | Range | 35 - 64 |
| Gender | Female | $n = 3$ |
| | Male | $n = 1$ |
| Length of time living in non-English speaking country (in years) | Mean | 7.5 |
| | Range | <1 - 25 |
| Hours/day speaking English with EAL learners | Mean | 5.0 |
| | Range | 3 - 6 |
| Experience as an examiner (in years) | Mean | 11.0 |
| | Range | 6 - 12+ |

Note. $N = 4$

The examiners' background questionnaire also included three open-ended questions: a) countries in which participants had worked as EAL instructors, b) countries in which participants had worked as EAL examiners, and c) examiners' thoughts about CL1 test-takers' performances. While the first question intended to establish examiners' exposure to cultures and languages other than English, the second question intended to collect information about their experience assessing

EAL speakers' speaking skills in countries other than their country of birth and about their exposure to different cultures and different languages. I considered these answers vital as they could shed a light on how the examiners' prior experiences might play a role on the assessment of the test-takers' performances. The last question intended to glean information about examiners' opinions and possible preconceptions about CL1 test-takers' performances in general. The answers to these questions can be summarized in Tables 2 and 3.

Table 2

Countries in Which Participants Worked as EAL Instructors and as Examiners

| Examiners | Countries (taught English) | Countries (worked as an examiner) |
|------------|--|--------------------------------------|
| Examiner A | Singapore, Indonesia, Japan, Canada, UK | Singapore, Canada |
| Examiner D | South Korea, Japan, South Africa, Canada | South Korea, Japan, Canada |
| Examiner E | Hong Kong, Canada | Hong Kong, Canada |
| Examiner F | Canada, Ecuador | Canada |

Note. $N = 4$

Table 3

Examiners' Questionnaire Answers - Thoughts about Chinese-as-a-first-language Test-takers' Performances

| Examiners | Thoughts |
|------------|--|
| Examiner A | "Differences between mainland and overseas Chinese"; "memorized"; "formulae" |
| Examiner D | "They tend to study for/take tests well and have trouble applying what they know beyond the test." |
| Examiner E | "I've noticed a tendency to memorize and regurgitate - sometimes entire paragraphs." |
| Examiner F | "I am occasionally challenged by their pronunciation." |

Note. $N = 4$

3.3.2 Test-takers' Background Questionnaire.

The test-takers' background questionnaire (Appendix E) collected information on their educational backgrounds, their experience as IELTS test-takers, the length of residence in their country of birth, the number of hours a day spent speaking English with fluent speakers, and their experience as second language learners and speakers of English.

All test-takers were female, between the ages of 18 and 44, and, at the time of the study, had been studying English between 11 and 25 years, with a mean of 18. Participants reported having lived in Canada between less than one year and 5 years. All test-takers had previously taken the IELTS Speaking Test⁵.

Table 4 provides information on the test-takers' characteristics.

⁵ The self-reported scores and information on training received for the IELTS were used for an initial comparison to the scores assigned to the test-takers in this study in an attempt to identify possible discrepancies.

Table 4

Test-takers' Characteristics

| | | |
|--|--------|--------------|
| Age (in years) | Mean | 29.9 |
| | Range | 18 - 44 |
| Gender | Female | <i>n</i> = 4 |
| Length of residence in country of birth (in years) | Mean | 22.7 |
| | Range | 16 - 25+ |
| Age test-taker started studying English | Mean | 13.9 |
| | Range | 10 - 20 |
| Length of time in Canada (in years) | Mean | 1.5 |
| | Range | <1 - 5 |
| Length of time as an English language learner (in years) | Mean | 18 |
| | Range | 11 - 25 |
| Hours/day speaking English with fluent speakers | Mean | 1.25 |
| | Range | <1 - 2 |
| Received training for the IELTS test | Yes | 2 |
| | No | 2 |
| Times the IELTS test was taken | Mean | 2 |
| | Range | 1 - 3 |
| Most recent score | Mean | 6.6 |
| | Range | 6.0 - 7.0 |

Note. *N* = 4

3.3.3 Modified Rating Sheet

To record scores assigned to each test-taker, examiners received a modified rating sheet (see Appendix F). The rating sheet was constructed to collect information on the scores assigned to each of the four criteria (*fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation*) assessed during the speaking test. The rating sheet also provided examiners with additional space to include information about their scoring processes and decisions.

3.3.4 The IELTS Speaking Test.

The speaking component of the IELTS, a three-part, one-to-one interaction between an examiner and a test-taker, has been designed to evaluate a test-taker's overall language knowledge and ability to communicate effectively with English language speakers (IELTS, 2016). The test takes between 11 to 14 minutes, and consists of three parts, each following a specific pattern of tasks to assess test-takers' speaking competence in different ways. The three parts are:

- Introduction and interview, in which the examiner asks general questions about the test-taker and familiar topics;
- Individual long turn, in which the test-taker is given a card with a topic and points to be included in his/her talk;
- Two-way discussion, in which the examiner asks questions designed to elicit answers that lead to discussion of abstract ideas and issues related to the topic presented in the individual long turn (IELTS, 2016).

A detailed description of the speaking test can be found in Table 5.

Table 5

Description - IELTS Speaking Test

| Section | Duration | Information |
|--|------------------|---|
| Part 1 Introduction and interview | 4 - 5 minutes | The examiner will introduce him or herself and ask you to introduce yourself and confirm your identity. The examiner will ask you general questions on familiar topics, e.g. home, family, work, studies, and interests. This section should help you relax and talk naturally. |
| Part 2 Individual long turn | 3 - 4 minutes | The examiner will give you a task card which asks you to talk about a particular topic, including points to include in your talk. You will be given one minute to prepare and make notes. You will then be asked to talk for 1 - 2 minutes on the topic. You will not be interrupted during this time, so it is important to keep talking. The examiner will then ask you one or two questions on the same topic. |
| Part 3 Two- way discussion | 4 - 5 minutes | The examiner will ask you further questions which are connected to the topic of Part 2. These questions are designed to give you an opportunity to discuss more abstract issues and ideas. |

Note. Information retrieved from IELTS.org.

Test-takers' speaking competence is assessed based on four criteria: *fluency and coherence* (speaking at length, organizing ideas logically), *lexical resource* (the vocabulary test-takers use, the appropriateness of words used to convey meaning, the organization of ideas in a logical order, and the use of cohesive devices), *grammatical range and accuracy* (the range and accuracy of the grammar test-takers use), and *pronunciation* (test-takers' ability to speak in an understandable way). Test-takers' performances are rated based on the IELTS band descriptions, which range from 1 (unable to communicate in English) to 9 (fluent speaker). Information on the band descriptors can be found in Appendix G.

Some important factors led to the selection of the IELTS Speaking Test as the standardized test to be used in this study. First, using the IELTS Speaking Test would minimize potential

issues that could arise from the use of a non-standardized test, such as questions regarding the validity of the instrument, while providing me with the opportunity to observe the nonverbal cues as they were being produced during the face-to-face exchanges. Second, unlike other standardized face-to-face speaking tests such as the Cambridge English: Advanced, the IELTS Speaking Test is a *one-on-one* interaction between an examiner and a test-taker. Using tests that consist of the participation of two test-takers could minimize examiners' ability to recall their perceptions of each individual test-taker's nonverbal cues. Moreover, the nonverbal cues produced between the EAL student/EAL student dyad might differ from those produced during exchanges between one EL1 examiner and one EAL test-taker. Finally, IELTS is one of the most recognized language proficiency tests in the world, being accepted in over 140 countries (IELTS, 2016). IELTS also continuously invests in research that investigates the reliability of scores assigned and the validity of its testing system (IELTS, 2016).

Being unable to have access to an official IELTS Speaking Test, sample tests were randomly selected. Initially, six different tests were chosen; however, to minimize limitations that might have resulted from ⁶differences in topics, one test was used in all speaking tests conducted in this study. The test was selected based on feedback elicited from participants who took part in the pilot study.⁷

3.4 Video-stimulated Verbal Recall Sessions

To learn about the examiners' perceptions of test-takers' nonverbal cues, I conducted video-stimulated verbal recall sessions with each examiner immediately after the conclusion of

⁶ The topic chosen might have influenced the test-takers' performances. Although environmental issues are a concern in the People's Republic of China, some test-takers might have felt uncomfortable criticizing the role of the government in dealing with those problems (T.-C. Lin, personal communication, December 20, 2016).

⁷ More information can be found in Section 3.6.

each face-to-face speaking test. I believed that having access to examiners' recalls of their thoughts could provide a rich account of their perception of test-takers' nonverbal cues and of their decision-making process.

Video-stimulated recall uses visual reminders "to prompt participants to recall thoughts they had while performing a task or participating in an event" (Gass & MacKey, 2000, p. 17) as it is believed that mental processes that take place during an event or task may be later recalled with the aid of a visual reminder. Stimulated recall sessions have been identified as a practical method for collection of data on cognitive processes, perspectives on learning, and interpersonal behaviour of participants (Mackey & Gass, 2005). Research has suggested that giving participants an opportunity to relive events may provide better access to their cognitive processes (Gass & Mackey, 2000).

Concerns in regards to the validity of the data collected during video-stimulated verbal recall sessions have been raised (Morthorst et al., 2012). One of the main criticisms relies on the uncertainty about the accuracy of self-reports due to the time lapse between tasks and verbal reports. Time lapse is often mentioned as a factor that may hinder the ability of participants to recall sessions (Huang, 2014). One of the main arguments is that, after long gaps between the event and the recall sessions, participants may recall what they *believe* they were thinking while the event was taking place rather than what their mental processing was. Despite criticism, video-stimulated verbal recall sessions have been commonly used as a method for observation of participants' cognitive processes (Huang, 2014). Immediate recall sessions (also called consecutive recalls), which take place immediately after the event, may provide researchers with the best opportunity to collect more reliable data (Gass & Mackey, 2000). Awareness that time lapse might compromise the reliability of the findings resulted in my decision to conduct the recall sessions with examiners immediately after each speaking test (Huang, 2013).

3.5 Pilot Study

In February 2015, two participants were recruited to take part in a pilot study, which aimed to assess the designed instruments for data collection and to identify possible unforeseen procedural problems that could negatively affect data collection. The participants, one international graduate student and one domestic graduate student who were both familiar with the IELTS Speaking Test, were assigned the roles of test-taker and examiner, respectively. During the pilot study, participants followed the procedures established for the main study. The examiner and the test-taker filled out background questionnaires before the speaking test began. The speaking test was conducted following the established protocols, and I conducted a video-stimulated verbal recall session with the examiner immediately after the conclusion of the speaking test.

Following the pilot study, I modified the background questionnaires and the script for the video stimulated verbal recall sessions. The changes in the questionnaires involved the revision of open-ended questions regarding age, number of years as EAL students, number of years as examiners, length of time living in the place of birth, and number of hours spent speaking with EAL speakers with number ranges (e.g., 1 to 5, 6 to 10, and so on). Refinements were also made to the video-stimulated verbal recall script; specific questions about the examiner's perception of the test-taker nonverbal cues were added to elicit more precise information. Below are some of the questions added/refined:

1. I noticed that you paused/looked confused/smiled/frowned. What were you thinking here?
2. I noticed you were smiling/frowning/nodding/looking at the test-taker. Can you tell me what your thoughts were?
3. Do you remember any thoughts regarding the test-taker's behaviour in this part?

4. After the conclusion of the recall session: Is there something you would like to add before we finish?

Finally, as both participants were graduate students familiar with the IELTS Speaking Test, I asked the participants to provide feedback on the sample speaking tests originally chosen for the study. Based on the opinions shared, one topic was selected for the main study (see Appendix A). Participants stated that their choice was due to two main factors: a) how current the topic was, and b) how relevant the topic was to all individuals regardless of cultural background, age, or gender.

3.6 Main Study

3.6.1 Recruitment

Following the proper procedures as approved by the University of Victoria Ethics Board, recruitment for the study started in February 2015. To recruit test-takers, I contacted graduate secretaries at the university in which the research took place and requested an invitation (see Appendix B), which was attached to the message, to be forwarded to all graduate students in every department. Recruitment of examiners was also done via email. I sent invitations (see Appendix C) to language schools and asked directors/managers to share my invitation with all the instructors in their institutions. Individuals contacted me via email to state their interest in taking part in the study.

As individuals contacted me, I confirmed if the criteria established for participation were met. Participation was voluntary. Test-takers received a \$10 Starbucks gift card and examiners financial compensation for their participation. Examiners' compensation was based on the hourly rate of C\$ 25.00. Speaking tests were scheduled based on participants' availability. Each examiner was assigned a letter and each test-taker a number.

3.6.2 Data Collection Procedures

The main study consisted of two data collection sessions: a) a face-to-face speaking test followed by a video-stimulated verbal recall session with an examiner and b) ratings of audio recordings of the face-to-face speaking tests.

Session one. The first data collection session consisted of six individual speaking tests, which took place at a research lab between March and April 2015, at the University of Victoria. During that session, examiners reviewed and signed consent forms (see Appendix H), which provided information about their participation, time commitment, and their role in the study. They were informed of their right to withdraw at any time and to ask questions if any information was found to be unclear. Once consent forms were signed, examiners filled out a background questionnaire (see Appendix D). Next, the examiners reviewed the speaking test, the modified rating sheet, the IELTS scoring rubric (see Appendices A, F, and G), and a sample script of an IELTS Speaking Test.

While the examiner reviewed the speaking test topic and questions, the test-taker was placed in a separate room to review and sign the consent form (see Appendix I), which provided information about participation, time commitment, and role in the study. All test-takers were also informed of their right to withdraw at any time. Upon signing the consent form, the test-taker filled out a background questionnaire (see Appendix E). Once the questionnaire was completed, the test-taker was invited to join the examiner in the room in which the speaking test took place. Both participants were then asked to treat the simulated test as a real IELTS Test. They were also reminded that my role during the speaking test would be that of an observer.

Next, one examiner conducted a face-to-face speaking test with one test-taker (e.g., Examiner A with Test-taker 1). The test-taker was seated across from the examiner. An iPad 2

was placed on the table at which participants were sitting to capture both visual and audio data. A Canon VIXIA HF R21 digital camcorder was placed on the left corner of the room facing both the examiner and the test-taker to capture a different angle of their interaction and to provide a backup of the video recording of the speaking test. To collect the audio recording of the speaking tests, an Olympus VN-8100PC digital voice recorder was placed on the table in front of participants. I sat behind the test-taker to prevent visual interference. Immediately after the completion of all three tasks that comprised each speaking test, a video-stimulated verbal recall session with the examiner took place. The session was recorded using an Olympus VN-8100PC digital voice recorder. Both the script and the protocol were followed. The video recording of the speaking test was played to the examiner on the iPad. The examiner was asked to focus on eye contact, smiles, and gestures, and to inform how those were perceived during the speaking test. The examiner was also asked to discuss the test-taker's performance and to expand on his/her notes and ratings. Both the researcher and the examiner decided when to stop the video and make comments (Huang, 2013).

Session two. In this session, which took place in April 2015, examiners rated audio recordings of all face-to-face speaking tests. Prior to sending the recordings to the examiners, test-takers were randomly assigned new numbers. Examiners were asked to follow the exact same protocols and rubric when assessing the audio-recorded performances. Audio-recordings of all the tests and new rating sheets (see Appendix J) were then emailed to each examiner. Scores were returned within seven days from the day they were sent to examiners. Information on data collection sessions can be found in Table 6.

Table 6

Data Collection Sessions

| Sessions | Examiner A | Examiner B | Examiner C | Examiner D | Examiner E | Examiner F |
|--|--|--|--|--|--|--|
| Session one (1 hour and 10 minutes) | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Review of IELTS scoring rubric and speaking test topic and questions (15 minutes)</p> <p>Face-to-face speaking test – candidate 1 (11 to 14 minutes)</p> <p>Audio-recorded video-stimulated verbal recall session (30 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Review of IELTS scoring rubric and speaking test topic and questions (15 minutes)</p> <p>Face-to-face speaking test – candidate 2 (11 to 14 minutes)</p> <p>Audio-recorded video-stimulated verbal recall session (30 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Review of IELTS scoring rubric and speaking test topic and questions (15 minutes)</p> <p>Face-to-face speaking test – candidate 3 (11 to 14 minutes)</p> <p>Audio-recorded video-stimulated verbal recall session (30 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Review of IELTS scoring rubric and speaking test topic and questions (15 minutes)</p> <p>Face-to-face speaking test – candidate 4 (11 to 14 minutes)</p> <p>Audio-recorded video-stimulated verbal recall session (30 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Review of IELTS scoring rubric and speaking test topic and questions (15 minutes)</p> <p>Face-to-face speaking test – candidate 5 (11 to 14 minutes)</p> <p>Audio-recorded video-stimulated verbal recall session (30 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Review of IELTS scoring rubric and speaking test topic and questions (15 minutes)</p> <p>Face-to-face speaking test – candidate 6 (11 to 14 minutes)</p> <p>Audio-recorded video-stimulated verbal recall session (30 minutes)</p> |
| Session two (1 hour) | <p>Session two ?</p> <p>Ratings of speaking test (audio recording) - test-takers 1, 2, 3, 4, 5, and 6</p> | <p>Session two ?</p> <p>Ratings of speaking test (audio recording) - test-takers 1, 2, 3, 4, 5, and 6</p> | <p>Session two ?</p> <p>Ratings of speaking test (audio recording) - test-takers 1, 2, 3, 4, 5, and 6</p> | <p>Session two ?</p> <p>Ratings of speaking test (audio recording) - test-takers 1, 2, 3, 4, 5, and 6</p> | <p>Session two ?</p> <p>Ratings of speaking test (audio recording) - test-takers 1, 2, 3, 4, 5, and 6</p> | <p>Session two ?</p> <p>Ratings of speaking test (audio recording) - test-takers 1, 2, 3, 4, 5, and 6</p> |
| Sessions | Test-taker 1 | Test-taker 2 | Test-taker 3 | Test-taker 4 | Test-taker 5 | Test-taker 6 |
| Session one (30 min) | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Face-to-face speaking test - examiner A (11 to 14 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Face-to-face speaking test - examiner B (11 to 14 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Face-to-face speaking test - examiner C (11 to 14 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Face-to-face speaking test - examiner D (11 to 14 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Face-to-face speaking test - examiner E (11 to 14 minutes)</p> | <p>Session one ??</p> <p>Informed consent (5 minutes)</p> <p>Background questionnaire (5 minutes)</p> <p>Face-to-face speaking test - examiner F (11 to 14 minutes)</p> |

3.7 Transcriptions

Before data analysis was conducted, the audio and video recordings from the two data collection sessions were renamed. 100% of the video and audio recordings were transcribed.

3.7.1 Transcription of verbal data

The audio recorded speaking test and the recall sessions were transcribed after the nonverbal cues were coded. The recall sessions were first transcribed into individual Microsoft Word 2011 documents. Next, I reviewed all the transcriptions to check for accuracy. To transcribe the face-to-face speaking tests, I used Express Scribe⁸, which allows for the speed of the audio recording to be increased or decreased as needed, to play the audio files while the transcriptions were entered in individual Microsoft Word 2011 files. The transcriptions were reviewed for accuracy one week after they were first concluded. No inconsistencies were found.

Upon conclusion of the all transcriptions, the nonverbal and verbal data sets were integrated as the complexity of interpersonal (and intercultural) communication makes the concurrent examination of both data sets an important component for the achievement of contextual understanding.

3.8 Coding

During the first coding session, I used NVivo 10 for Mac. All data were initially entered and coded. Limitations to the software's ability to perform essential tasks that are available in its Windows version, such as the creation of theme trees and concomitant playback and coding of video data (which would allow for inclusion of timeline in the coding of nonverbal cues), resulted

⁸ Express Scribe is professional audio player software used to assist the transcription of audio files.

in my decision to conduct the following coding sessions manually.⁹

3.8.1 Nonverbal cues

To ensure that consistency in the coding of the nonverbal cues would be achieved, a coding scheme (see Appendix K) was adapted from the Bernese System for Coding Nonverbal Interactions (Hirsbrunner, Frey, & Crawford, 1987), created to code cues produced during dyadic exchanges in which interlocutors are seated. Nonverbal cues were manually coded. The need to adapt the coding scheme for this study took into consideration some important factors:

- The Bernese System does not provide information on coding of movements produced by one interlocutor towards another (i.e., movements that indicate interaction between interlocutors) but rather defines dimensions (e.g., up/down shift of hand, angle of hand in vertical plane). It also does not provide codes for gaze direction;
- The temporal component of the data was considered vital as it would allow for the analysis of sequences of production of the three nonverbal cues this study aimed to investigate. The information about the moment the cues were produced also allowed for comparison of moment-by-moment production of cues by both the examiners and test-takers; and
- The coding scheme allowed for the systematic and consistent coding of the three nonverbal cues produced during the exchanges.
- The Bernese System provides a coding scheme has been shown to be consistent (Hirsbrunner, Frey, & Crawford, 1987);

I started coding the observable nonverbal cues produced by both examiners and test-takers

⁹ All the coding procedures described henceforth refer to manual coding of the data.

after all the data were collected. To code the nonverbal data, I used QuickTime for Mac to display each individual frame (10 frames per second) of every video clip¹⁰. The coding was done frame-by-frame and entered in a Microsoft Excel 2011 sheet that was designed to include timing and observable nonverbal cues produced by both the examiner and the test-taker. The time of production of nonverbal cues was entered in second-by-second increments. The sheet was divided into two sections: examiner's and test-taker's. A sample of the transcription of the nonverbal data can be found in Table 7.

¹⁰ Although some video systems allow for an analysis of up to 60 frames/s or more, other studies have also shown that longer intervals (such as half-minute) also provide adequate data on sequences of movements (Hirsbrunner, Frey, & Crawford, 1987).

Table 7

Sample of an Ongoing Coding Session of Nonverbal Data (Examiner F and Test-taker 6)

| Examiner F | Nonverbal Cues | Test-taker 6 | Nonverbal Cues |
|-------------------|-----------------------|---------------------|--|
| 0:07:02 | | 0:07:02 | Opens right hand; stretches right arm out ("the population of..."); looks straight ahead (away from examiner) |
| 0:07:03 | Eye contact | 0:07:03 | Eye contact; grips pen with left hand |
| 0:07:04 | | 0:07:04 | Looks straight ahead (away from examiner) |
| 0:07:05 | Eye contact | 0:07:05 | Eye contact |
| 0:07:06 | Eye contact | 0:07:06 | Looks to the right (away from examiner); extends right arm; brings right arm in; moves right hand ("four thousand million"); eye contact |
| 0:07:07 | | 0:07:07 | Moves right hand up once; laughter; Duchenne smile |
| 0:07:08 | Non-Duchenne smile | 0:07:08 | Moves pen to the left hand; looks straight ahead (away from examiner) |
| 0:07:09 | | 0:07:09 | |
| 0:07:10 | Eye contact | 0:07:10 | Moves both hands up and down slightly (palms facing each other) ("I think so"); eye contact; looks straight ahead (away from examiner) |
| 0:07:11 | | 0:07:11 | |
| 0:07:12 | | 0:07:12 | Moves right hand back and forth; moves both hands to the right in an up-down motion ("strong impact on people") |
| 0:07:13 | Eye contact | 0:07:13 | Eye contact |
| 0:07:14 | | 0:07:14 | Moves hands up and down (palms open) ("living there") |
| 0:07:15 | | 0:07:15 | Looks straight ahead (away from examiner); grips pen with the right hand |
| 0:07:16 | | 0:07:16 | Looks down (away from examiner); opens right hand (palm up) ("And also...") |
| 0:07:17 | | 0:07:17 | Grips pen with right hand |
| 0:07:18 | | 0:07:18 | Moves right hand up and down (from right to left) ("Maybe it is...") |

On the first coding session, all observable nonverbal cues produced by both the examiner and the test-taker in each videotaped dyadic interaction were coded separately. The nonverbal

cues were coded as follows: (1) smiles; (2) eye contact; (3) hand gestures and arm movements; (4) body position; and (5) head position. The first stage resulted, therefore, in five initial transcripts per participant, which were later condensed into one single script per participant. The 50 minutes and 20 seconds of video recording that comprised the four clips were coded. The first stage of coding took approximately 8 weeks.

During the second stage, all nonverbal cues that were not pertinent to this study were omitted. The remaining nonverbal cues were then coded into three broader categories: gestures, smiles, and eye contact. Each category was coded separately, resulting in three coding sheets for each participant, which were later combined into a single document per participant¹¹. Next, nonverbal cues from each dyadic interaction (examiner and respective test-taker) were entered in one single page. Upon conclusion of this stage of coding, the first and the second sets of codes were compared second-by-second to check for accuracy. Despite the fact that transcripts were done in ten frames per second, review of the recordings indicated that some cues had not been identified in the first stage. The cues that had not been originally identified were: one smile produced by test-taker 6 and three episodes of eye contact (one from test-taker 1 and two from test-taker 6). To verify the accuracy of the second round of coding, the video clips and codes were reviewed.

In the last session, two out of the three broad categories (smiles and gestures) were broken down into subcategories¹². Smiles were discriminated as either Duchenne or non-Duchenne, and gestures were classified as illustrators and adaptors (self- and object-). The exclusion of emblematic and regulators was due to the following factors:

¹¹ Although it would have been possible to enter all codes in a single document during the first coding session, concerns about my ability to maintain consistency throughout all the coding sessions led to the decision to focus on one specific cue per participant per session.

¹² Eye contact remained as a single category.

- Both consist not only of hand gestures but also of other cues, such as head nods and posture shift, which were beyond the scope of this study;
- Emblematics transmit messages that are understood by most individuals from the same subculture (Ekman & Friesen, 1977). Inclusion of those gestures would invariably result in limitations regarding my ability to decode the data as my cultural background differs from both examiners' and test-takers'; and
- Emblematics and regulators can be consciously produced. Inclusion of those gestures would require variables beyond the scope of this study to be taken into account, including examiners' awareness of test-takers' intention to communicate using a specific gesture.

The decision to examine illustrators and adaptors as separate categories was the result of the following factors:

- Illustrators have been linked to the semantic and pragmatic functions of speech and are considered fundamental to human communication (Holler & Wilkin, 2011; Kendon, 2000, 2004). Moreover, illustrators may contribute to the information contained in a speaker's message (Holler et al., 2009);
- Adaptors (self- and object-), on the other hand, have been connected to feelings of nervousness and uneasiness (Lewis et al., 2015). Adaptors have also been linked to avoidance strategies (Canry & Spitzberg, 1987).

As illustrators and object- and self-adaptors perform different functions in face-to-face exchanges, the investigation of each individual type was considered valuable for the proper assessment of their role in how examiners perceived and assessed test-takers' performances. Figures 2, 3, and 4 illustrate the three gestures investigated.

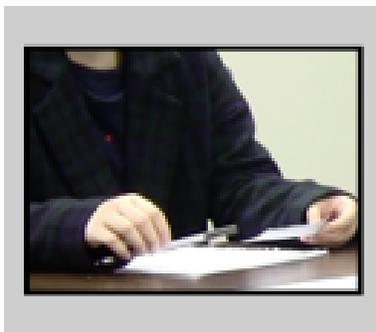


Figure 2. Example of object-adaptor.

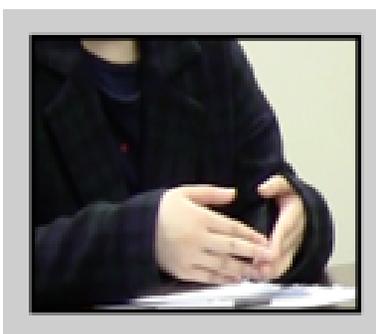


Figure 3. Example of self-adaptor.

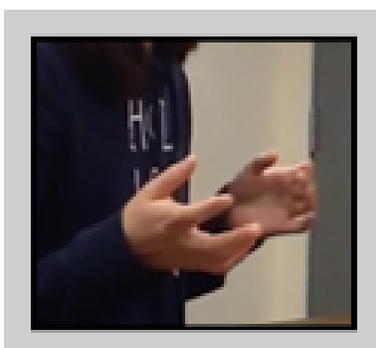


Figure 4. Example of illustrator.

Eye contact was calculated in terms of frequency of production¹³. Although measurement

¹³ Eye contact frequency refers to the number of glances made both examiner and test-taker.

of the length might provide information regarding the total amount of time that test-takers and examiners maintained eye contact, the decision to look at frequency resulted from the lack of equipment that could accurately calculate the length with precision.

Next, gestures and smiles were coded for production alongside speech or production in isolation (i.e., when test-takers were not speaking). Findings from the analysis could not only contextualize the use of the nonverbal cues but also provide more precise information on the possible influence of each cue on the overall interaction between examiners and test-takers. Moreover, investigating cues that were produced more frequently while test-takers spoke or were silent could lead to a better understanding of how examiners perceived the cues generated and their perception of test-takers.

During the last coding session, the observable nonverbal cues produced by test-takers were also identified for mimicry. This round was completed within a week after completion of the second coding session. 100% of the transcribed data were coded. The coding of the three nonverbal cues was conducted twice, one week apart.

I used the codes and the video recordings to identify possible mimicry of nonverbal cues. Nonverbal cues that were mimicked by examiners were tagged. The decision to investigate mimicry was due to the fact that:

1. Implicit behavioural mimicry has been shown to regulate social interactions (Hall, Millings, & Bouças, 2012).
2. Mimicking has been shown to help foster relationships and increase rapport in social interactions (Chartrand & Bargh, 1999; Lakin et al., 2003, 2008) and liking between interlocutors (Chartrand & Bargh, 1999).
3. Analysis of mimicked nonverbal cues might provide an insight into their role in how examiners perceived test-takers' overall behaviour and how they assessed their

performances during the face-to-face speaking tests.

As the aim of this study was to investigate the role of the nonverbal cues *produced by test-takers* on scores assigned by examiners, the number of cues initiated by the examiners that were mimicked by the test-takers was not included in the data analysis.

Finally, every transcript was reviewed twice to verify their accuracy and consistency.

3.8.2 Video stimulated verbal recall sessions.

The video-stimulated verbal recall sessions were coded for identification of patterns once the coding of the nonverbal cues was finalized. Prior to the beginning of the first coding stage, the data were labeled to protect participants' anonymity. Next, I familiarized myself with the data, listening to the recordings of the recall sessions and reading the transcripts several times. This process gave me the opportunity to document initial impressions (field notes) about examiners' responses. Next, to create codes¹⁴, I used the impressions, ideas, and notes that were produced during the initial stage and segments that contained patterns that I considered of interest. All codes included extracts of the data to provide contextual information for the creation of themes, which were generated in the next stage. I coded all the transcripts and organized the codes into groups.

In the next stage, the codes and respective extracts were organized to create possible themes based on elements that could be grouped into broader units of analysis (Boyatzis, 1998). Some of the initial codes, such as *examiner's self-perception*, were discarded during this stage as they were beyond the scope of this study. Initial themes that were identified included perceptions of test-takers' verbal and nonverbal behaviour, examiners' use of nonverbal cues to convey

¹⁴ Here, code refers to "the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon" (Boyatzis, 1998, p. 63).

messages during speaking tests, their perception of IELTS examiners, and their feelings during the test. The codes were initially organized into themes and sub-themes (e.g., *perception of test-taker's behaviour* -> *more/less confident*) based on how encompassing they were and how they related to each other (Braun & Clarke, 2006). An example of the organization of some of the initial themes can be found in Figure 5.



Figure 5. An example of the organization of the themes.

During the next stage, I reviewed the extracts, the codes, and the generated themes. It is important to note that the final themes do not reflect the frequency in which they were identified in the data (i.e., that they were the most frequent) but rather key topics, which, together, contained important elements that could help me answer my research question. The decision on which themes to discard and which to keep, therefore, was based on my personal judgment and

on their relevance to the study. For instance, one examiner stated relating to the test-takers' personal experiences, which were shared during the speaking test (throughout this recall session, this feature was identified nine times); however, this topic was beyond the scope of this study, resulting in my decision to remove it from the data analysis. The five codes used in the coding of the data can be found in Table 8. Overall, the codes were identified 46 times. Table 8 includes the themes identified in the analysis, the codes created for the analysis of the data, and examples of the notes that were taken throughout the data collection stage. These notes were later used to guide my analysis as they offered me an opportunity to reflect on my first impressions.

Table 8

Video-stimulated Verbal Recall Sessions - Themes, Codes, Examples, and Observations

| Theme and Definition | Code | Example | Notes | Frequency |
|--|------------------------------------|---|---|-----------|
| <p>Perception of test-taker's smiles Examiners' thoughts about smiles produced by the test-takers during the speaking tests</p> | Test-takers' smiles | "[. . .] and I think the other sort of smile was that she talked about something personal instead of something abstract. So she started "ooh it's my birthday," and she kind of made a joke, right?" | Did the examiner associate the test-taker's smiles to personal stories shared during the test? Did other examiners do the same? | 6 |
| <p>Perception of test-taker's use of gestures Examiners' thoughts about the gestures produced and used by the test-takers during the speaking tests</p> | Test-takers' gestures | "I've never seen somebody do that (hand gesture) before. I've seen people do other things when they're looking for a particular word. They probably don't even know they do, but at that point in time, yeah... I sort of was somewhat distracted by that." | Did the hand gestures produced by the test-taker affect the scores assigned to her? Look at frequency of production. She did produce a lot of hand gestures and smiles. | 7 |
| <p>Perception of test-taker's eye contact Examiners' thoughts about test-takers' eye contact during the speaking tests</p> | Test-takers' eye contact | "In the first section, I feel her eye contact with me was natural and logical and not noticeable in any way. It was a Canadian style, or it was a test-taking style; there wasn't anything special" | Did the examiner mean that there is a different eye contact for non-Canadians? What would that be? Did the examiner feel closer to the test-taker because of that? | 13 |
| <p>Perception of test-taker's behaviour Examiners' thoughts about test-takers' overall behaviour during the speaking tests</p> | Test-takers' behaviour | "Her behaviour seems really natural to me for a student who's in a slightly uncomfortable situation." | Look into this recall session and verify how frequently the examiner refers to things being 'natural' and a possible positive attitude towards the test-taker. | 7 |
| <p>Encouragement through nonverbal cues Examiners' encouragement using nonverbal rather than verbal communication during the speaking tests</p> | Examiners' nonverbal encouragement | " "Cause you can't say anything in an IELTS test [. . .] so encouragement has to be nonverbal. And I think if you smile at somebody or nod, you know...it gives cues that you're listening and paying attention..." | See if other examiners mention using nonverbal cues to encourage test-takers. Do they all mention IELTS, too, or are they more general? | 13 |

Next, a second coder independently coded 50% of randomly selected data. The second coder was an experienced researcher who had previous experience as an adult EAL instructor. The calculated inter-coder reliability was 94.74%. One discrepancy was found in the coding of the following segment of examiner E's recall session:

- “And I was happy that she came in smiling, ready to... to look at me, to talk to me.”

While I initially coded this extract as *perception of test-taker's behaviour*, the second coder coded it as *perception of test-taker's smiles*.

- The second coder and I discussed the reasons for our choice of coding. While I understood the statement as the examiners' perception of the test-taker's overall positive attitude in a broader context (“ready to look at me, talk to me”) rather than just being an observation about the test-taker's smile, the second coder perceived it as the examiner's highlighting the test-taker's smiles and the examiner's reaction to it. Upon discussion about the different possible interpretations for the excerpt, it was decided that it should be coded as *examiner's perception of test-taker's smiles*.

3.9 Speaking Tests

Although this study aimed to investigate the role of nonverbal cues on scores assigned to EAL test-takers, I believed that examining each test-taker's performance could provide important information regarding the examiners' scoring process. Analysis of the performances might also enrich the study; cross-referencing with the frequency of production of nonverbal cues could provide valuable information on how scores were assigned.

As previously discussed, the IELTS Speaking Test consists in the assessment of four criteria: a) *fluency and coherence*; b) *lexical resource*; c) *grammatical range and accuracy*; and d) *pronunciation*. Although the study might benefit from the inclusion of the analysis of test-

takers' performance in each criterion, features that are assessed in three out of the four criteria precluded me from performing a rigorous, consistent, and objective measurement. The rationale for the exclusion of criteria are as follows:

1) *Fluency and coherence*: aspects of both *fluency and coherence* that are assessed during the IELTS Speaking Test include length of speaking, logical organization of ideas, and appropriate indication of thought directions¹⁵. Two elements presented challenges in how the data could undergo a rigorous and objective analysis. Hesitation/speech rate have been calculated in previous studies in terms of ratio of filled and unfilled pauses to speech measured in milliseconds. Time and resource constraints (lack of specific software) prevented a rigorous analysis of this feature from being conducted. Furthermore, questions regarding the measurement of response length ("speaks at length", Bands 6 and 7) remain as IELTS does not provide explicit information on the assessment of length (Brown, 2006b; Seedhouse & Harris, 2014). The decision, therefore, was to exclude this criterion from the analysis.

2) *Pronunciation*: the assessment of *pronunciation* consists of elements such as word and sentence stress and intonation. Review of the literature on L2 speaking skill assessments shows possible limitations to the conduction of a rigorous analysis of scores assigned to pronunciation skills. First, studies have indicated the possible subjectivity of the assessment of this criterion as the scoring rubric fails to indicate how objective measurements can be applied to test-takers' performances. Second, there is a lack of clear description of test-takers' pronunciation skills are evaluated (e.g., "uses a full range of pronunciation features with precision and subtlety"), which may lead to subjective assessments. Lastly, the analysis of this criterion would require better

¹⁵ Information regarding the aspects that are assessed in the four criteria was retrieved from the standardized information and video provided by British Council on its websites (refer to <https://www.britishcouncil.vn/en/exam/ielts/courses-resources/videos/grammar>).

audio recording devices. Due to these factors, *pronunciation* was not included in the analysis of the verbal data.

3) *Grammatical range and accuracy*: some of the aspects assessed during the speaking test include the balance between range and accuracy and the use of complex sentences. No clear explanation is provided by IELTS in regards to how this criterion is assessed, whether in terms of specific errors or global accuracy. Although some studies have found a significant relationship between grammatical errors and ratings, results have not always been linear (Iwashita et al., 2008). Studies have also indicated that grammar may be assessed implicitly during the assessment of language skills (Alderson & Banerjee, 2002). Finally, a definition of how test-takers' ability to "consistently (produce) accurate structures apart from 'slips' characteristic of native speaker speech" (band 9) (IELTS, 2016) can be objectively measured is not provided by IELTS in the band descriptors. As a result, the criterion *grammatical range and accuracy* was not included in the analysis.

As a result of limitations to the objective and reliable measurements mentioned above, the analysis of the four test-takers' verbal performance focused on an examination of test-takers' *lexical range*. The features assessed in this criterion are: a) test-takers' ability to paraphrase; b) test-takers' accurate and natural use of use of idiomatic language; and c) lexical breadth. Review of the literature provided no reliable measurements of L2 speakers' ability to paraphrase during a speaking test, a factor that has resulted in questions about the possible subjectivity involved in the assessment of this feature (Brown, 2002, 2006b). In terms of the assessment of how "naturally" test-takers use idiomatic language, the IELTS scoring rubric does not define "naturally" or provide information on how that could be reliably measured. Being unable to conduct an objective and rigorous analysis of these features, the study focused on test-takers' breadth of vocabulary.

Despite the fact that the assessment of this criterion includes evaluation of features such as the use of *a wide vocabulary resource readily and flexibly to convey precise meaning*” (IELTS, 2016), studies have shown that lexical range is a fundamental element of spoken fluency of L2 learners (Hilton, 2008; Kiozumi & In’ nami, 2013; Mewald, 2015). Research has also indicated a possible connection between lexical range and communicative competence and proficiency in speaking skills, suggesting that a broader lexical range might be linked to better spoken performance (Kiozumi & In’ nami, 2013; Mewald, 2015). To ensure a systematic measurement, copies of the transcripts containing only the test-takers’ performances were run through Lextutor’s¹⁶ vocabulary profile program. Lextutor is a text analysis tool widely used by researchers (Cobb & Horst, 2011; Gaskell & Cobb, 2004; Sorell & Shin, 2007). The report produced by the software provided information on the percentage of the text that fell within the most common 1,000 words in English (K1), the second most frequent 1,000 words (K2), words that pertain to the academic word list (AWL), and the percentage of words that could not be classified in the other three categories (Off-list).

3.10 Field Notes

Throughout the data collection sessions, I took notes and wrote down observations about features of the observable nonverbal cues that caught my attention during the face-to-face speaking tests. These notes were later used to inform the coding of the data.

For instance, in the notes taken during the fourth speaking test, I wrote down initial impressions in regards to the participants’ eye contact as it seems that test-taker 4 and examiner D did not look away from each other as often as test-taker 1 and examiner A had in the first

¹⁶ Detailed information about all the features available to researchers can be found on <http://www.lexutor.ca>.

speaking test; in fact, they seemed very comfortable maintaining eye contact. I compared my perception with the thoughts shared by the examiner about the test-taker's eye contact and to the frequency of production of that cue during the test. Results indicated that, indeed, test-taker 4 and examiner D did maintain eye contact for longer periods of time when compared to the first dyad.

Another example would be the notes I took on what seemed to be a very high frequency of nonverbal cues being generated by test-taker 6 during the speaking test; I wondered if they had been noticed by the examiner. I incorporated those notes to the analysis of examiner F's recall session data and later to the analysis of the nonverbal data and scores assigned.

Methodological concerns underpinned my decision to not incorporate some of the observations and notes taken during the data collection stage to the data analysis as, although valuable, they did not always provide accurate observations of the data that were collected. For instance, I had the impression that test-taker 1 was very reserved in terms of production of nonverbal cues, and my initial thought was that the test-taker would have produced the least number of cues when compared to the other three test-takers. Revision of the final coding showed, however, that this impression was inaccurate.

Although not all field notes were found to be relevant, repeated viewing of the videos provided me with an opportunity to revisit my initial observations and examine them against the recordings and re-evaluate observations that did not match findings from the data analysis.

3.11 Data Analysis

In order to address my research question, the data collected underwent qualitative and quantitative analyses, and the different sources of data were integrated. Being an exploratory study, the analysis of the data focused on investigating the possible role that the three nonverbal cues could have on the scores assigned during the face-to-face speaking tests and also on

developing an understanding of the complex phenomenon of intercultural communication within standardized speaking-test context.

Qualitatively, the video-stimulated recall and open-ended questions from the examiners' questionnaire underwent an inductive thematic analysis to explore examiners' thoughts and decision-making processes during the tests as well as their opinions of CL1 test-takers' nonverbal cues. Besides, the thoughts shared in the open-ended questions in the questionnaire presented an opportunity for the analysis of examiners' perceptions of CL1 test-takers' performances. The qualitative data were also examined vis-à-vis the scores assigned to test-takers' performances to gain a better understanding of their possible role in the examiners' assessments of test-takers' performances.

Intra-rater reliability was calculated. Scores assigned by each examiner to the face-to-face and audio-recording of the same test-taker (i.e., examiner A, test-taker 1; examiner D, test-taker 4; examiner E, test-taker 5; examiner F, test-taker 6) were entered in an Excel spreadsheet. Calculation using binary values (0 for agreement, and 1 for disagreement) did not provide the information needed for this study. To overcome that issue, calculations were conducted as follows¹⁷:

- a. Based on the IELTS scoring system, test-takers can be assigned scores from 1 to 9 (in 0.5 point increments), for a total of 17 (1.0, 1.5, 2.0, 2.5);
- b. Differences in scores assigned to the two sets of scores (audio-recorded and face-to-face) were calculated. A difference of 0.5 reflected one point, 1.0 reflected two points, and so on;
- c. The sum was deducted from the maximum of 17 points possible;

¹⁷ Calculation per test-taker.

- d. Results were divided by the total of 17 to achieve the percentage of intra-rater reliability.

The quantitative analysis was conducted as follows:

1. Test scores: scores were analyzed individually (scores assigned to each test-taker), holistically (overall scores assigned to all test takers), and analytically (scores assigned by criterion) (Huang, 2011);
2. The mean and the median of the test scores assigned to test-takers' audio-recorded performances were calculated and results were compared to the test scores assigned to the face-to-face speaking tests;¹⁸
3. Lexical range: the vocabulary used by test-takers during the speaking test was analyzed using Lextutor's vocabulary profile program. Results were analyzed vis-à-vis the scores assigned to the criterion *lexical resource*.
4. Frequency of production of nonverbal cues: eye contact, gestures, and smiles were analyzed holistically (total number of cues produced by all test-takers), individually (number of each individual cue produced by each test-taker), and by sub-categories (gestures – illustrators; self- and object- adaptors; smiles – Duchenne and non-Duchenne);
5. Production of nonverbal cues: nonverbal cues were analyzed based on their production alongside speech or in isolation;
6. Nonverbal cues and test scores: results from the analyses of nonverbal cues were analyzed vis-à-vis individual, analytical and holistic scores;

¹⁸ Although the use of the mean and median of the scores assigned to the audio-recorded performances could lead to questions regarding possible differences in how examiners assess test-takers' performances, the use of scores assigned to the audio-recorded tests to evaluate the reliability of the results follows the IELTS protocol. Test-takers who disagree with the scores assigned to their performances may request them to be re-assessed; in that case, the audio recording of the test is evaluated by a different examiner, and results are compared.

7. Mimicry: nonverbal cues generated by test-takers that were mimicked by examiners were analyzed vis-à-vis the test scores assigned.

Chapter 4: Results

This chapter presents the results of the methods used to analyze the data. Section 1 presents the analysis of the scores assigned to test-takers' performances in both the face-to-face and the audio-recorded speaking tests. Section 2 focuses on the results of the lexical analysis of the test-takers' performances, while Section 3 includes the results of the thematic and nonverbal data analyses.

4.1. Test Scores

The results I present here refer to the test scores assigned to the four test-takers' performances both in the face-to-face and in the audio-recorded speaking tests. Intra-rater reliability was calculated. The agreement rates were: examiner A, 88%; examiner D, 76%, examiner E, 88%, and examiner F, 82%.

An examination of the scores assigned to each test-taker's performance (face-to-face and audio-recorded) by the same examiner showed differences in the assessment of some of the criteria such as *fluency and coherence*, in which three out of the four scores assigned to the face-to-face performances were higher than those assigned to the audio-recorded ones (see Table 9).

Table 9

Face-to-face and Audio-recorded Performances (per Test-taker)

| Examiner A - Test- taker 1 | F | L | G | P | Final Score | Examiner D - Test- taker 4 | F | L | G | P | Final Score |
|--|----------|----------|----------|----------|------------------------|--|----------|----------|----------|----------|------------------------|
| Face-to- face speaking test | 7.0 | 7.0 | 6.0 | 6.0 | 6.5 | Face-to- face speaking test | 7.0 | 6.0 | 6.0 | 7.5 | 6.5 |
| Audio- recorded speaking test | 7.0 | 6.0 | 6.0 | 6.0 | 6.0 | Audio- recorded speaking test | 6.5 | 7.0 | 6.5 | 7.5 | 7.0 |
| Examiner E - Test- taker 5 | F | L | G | P | Final Score | Examiner F - Test- taker 6 | F | L | G | P | Final Score |
| Face-to- face speaking test | 7.5 | 7.5 | 7.0 | 8.0 | 7.5 | Face-to- face speaking test | 7.0 | 6.5 | 6.0 | 5.0 | 6.0 |
| Audio- recorded speaking test | 7.0 | 7.5 | 7.0 | 7.5 | 7.0 | Audio- recorded speaking test | 6.5 | 7.0 | 6.0 | 5.5 | 6.0 |

Next, to address the possibility of examiners being able to recall the scores they had assigned to the face-to-face performances, I examined the face-to-face scores against the scores assigned by all examiners to each audio-recorded performance. First, I calculated the overall mean scores of the face-to-face speaking tests and of the audio-recorded tests. These results were used to identify discrepancies in the scores assigned to the face-to-face and audio-recorded performances. The mean overall score assigned to all the test-takers in the face-to-face speaking tests was 6.69 while the mean score assigned to the audio-recorded tests was 6.52, a difference of

0.14.

The initial calculation of the overall scores did not account for how examiners evaluated the individual criteria (*fluency and coherence, lexical resource, grammatical range and accuracy, and pronunciation*) during the test. To address this, I analyzed the scores assigned to each criterion. Results showed that test-taker 5 received the highest scores in the face-to-face speaking test while test-taker 6 scored the lowest. Table 10 illustrates the mean scores and standard deviation by test-taker, and Table 11 presents the mean scores by criterion

Table 10

Audio Recorded Tests - Mean Scores and Standard Deviation by Criterion

| Test-taker | F | SD | L | SD | G | SD | P | SD |
|--------------|------|------|------|------|------|------|------|------|
| Test-taker 1 | 6.25 | 0.87 | 6.25 | 0.29 | 6 | 0.25 | 5.5 | 0.91 |
| Test-taker 4 | 6.5 | 0.70 | 6.5 | 0.48 | 6.5 | 0.25 | 6.75 | 0.64 |
| Test-taker 5 | 6.75 | 0.5 | 6.75 | 0.64 | 6.75 | 0.5 | 6.75 | 0.87 |
| Test-taker 6 | 6.5 | 0.25 | 7 | 0 | 6.25 | 0.29 | 5.5 | 0.29 |

Note. F - Fluency and Coherence; L - Lexical Resource; G - Grammatical Range and Accuracy; P - Pronunciation.

Table 11

Mean Scores Assigned by all Examiners to Audio-recorded Tests – Scores by Criterion

| Test-taker 1 | F | L | G | P | Final Score | Test-taker 4 | F | L | G | P | Final Score |
|-------------------------------------|----------|----------|----------|----------|--------------------|-------------------------------------|----------|----------|----------|----------|--------------------|
| Face-to-face speaking test | 7.0 | 7.0 | 6.0 | 6.0 | 6.5 | Face-to-face speaking test | 7.0 | 6.0 | 6.0 | 7.5 | 6.5 |
| Audio-recorded speaking test | 6.25 | 6.25 | 6.0 | 5.5 | 6.0 | Audio-recorded speaking test | 6.5 | 6.5 | 6.5 | 6.75 | 6.56 |
| Test-taker 5 | F | L | G | P | Final Score | Test-taker 6 | F | L | G | P | Final Score |
| Face-to-face speaking test | 7.5 | 7.5 | 7.0 | 8.0 | 7.5 | Face-to-face speaking test | 7.0 | 6.5 | 6.0 | 5.0 | 6.0 |
| Audio-recorded speaking test | 6.75 | 6.75 | 6.75 | 6.75 | 6.75 | Audio-recorded speaking test | 6.5 | 7.0 | 6.25 | 5.5 | 6.31 |

Note. F - Fluency and Coherence; L - Lexical Resource; G - Grammatical Range and Accuracy; P - Pronunciation.

As Table 11 shows, the scores examiners assigned to each criterion differed in the assessment of test-takers' performances in the face-to-face and the audio-recording of the speaking tests. In the assessment of *fluency and coherence*, all four test-takers received higher scores during the face-to-face speaking test when compared to those assigned to the mean of the audio recorded scores. Differences ranged from 0.5¹⁹ (test-takers 4 and 6) to 0.75 (test-takers 1

¹⁹ A difference of 0.5 is considered acceptable by IELTS in the case of a test being reviewed by a second examiner. That difference does not affect the overall assessment of test-takers' speaking skills.

and 5). An analysis of the criterion *pronunciation* showed similar results: three out of four test-takers received higher scores in the face-to-face speaking tests (with the exception of test-taker 6). In terms of *lexical resource* and *grammatical range and accuracy*, findings were inconclusive: test-takers 4 and 6 received lower scores during the face-to-face speaking while the scores from test-takers 1 and 5 showed that examiners assigned lower scores to their performances in the audio recorded speaking tests.

To conduct a more in-depth analysis of the scores, I calculated the median of the scores assigned by all examiners to each test-taker's performance in the audio-recorded speaking test and then compared these to the scores assigned during the face-to-face speaking tests. The analysis of the median scores addressed possible concerns about extreme high or low scores (outliers), which could have skewed the calculation of the mean scores. The results of the analysis are illustrated in Table 12.

Table 12

Median Scores by Criterion

| Test-taker 1 | F | L | G | P | Final Score | Test-taker 4 | F | L | G | P | Final Score |
|------------------------------|------|------|-----|-----|----------------|------------------------------|-----|-----|------|------|----------------|
| Face-to-face speaking test | 7.0 | 7.0 | 6.0 | 6.0 | 6.5 | Face-to-face speaking test | 7.0 | 6.0 | 6.0 | 7.5 | 6.5 |
| Audio-recorded speaking test | 6.25 | 6.25 | 6.0 | 5.5 | 6.0 | Audio-recorded speaking test | 6.5 | 6.5 | 6.5 | 6.75 | 6.56 |
| Test-taker 5 | F | L | G | P | Final Score | Test-taker 6 | F | L | G | P | Final Score |
| Face-to-face speaking test | 7.5 | 7.5 | 7.0 | 8.0 | 7.5 | Face-to-face speaking test | 7.0 | 6.5 | 6.0 | 5.0 | 6.0 |
| Audio-recorded speaking test | 7.0 | 6.75 | 7.0 | 7.0 | 6.94 | Audio-recorded speaking test | 6.5 | 7.0 | 6.25 | 5.75 | 6.37 |

Note. F - Fluency and Coherence; L - Lexical Resource; G - Grammatical Range and Accuracy; P - Pronunciation.

A comparison of the results in Tables 11 and 12 revealed no major differences in the scores assigned; however, the scores assigned to the criteria *fluency and coherence*, *lexical range*, and *pronunciation* presented consistent patterns in both sets of scores. In regards to *grammatical range*, differences were found in the assessment of test-taker 5's performance as the median scores, unlike the mean, did not indicate any differences between the scores assigned to the face-to-face and the audio-recorded tests.

4.2. Lexical Analysis

Following the analysis of the scores assigned to the criterion *lexical resource*, I investigated the lexical range of the performances of the test-takers. To examine the assessment of the test-takers' lexical range against the scores assigned, the vocabulary produced during the speaking tests was run through Lextutor. Results revealed that test-taker 5's percentage of K1 words is the highest of all test-takers (89.47%). Test-taker 1 produced the most K2 words (6.40%) while test-taker 6's frequencies of AWL and off-list words were the highest among the test-takers (4.15% and 6.84%, respectively). Next, I examined the number of word families, types, and tokens in each category. Table 13 shows the number of tokens per category.

Table 13

K-words - Families, Types, and Tokens

| K- words | | Test-taker 1 | Test-taker 4 | Test-taker 5 | Test-taker 6 |
|-----------------------------------|----------|--------------|--------------|--------------|--------------|
| K1 words (1-1000) | Families | 178 | 187 | 195 | 188 |
| | Types | 213 | 226 | 239 | 220 |
| | Tokens | 756 | 727 | 1011 | 823 |
| | % | 84.85 | 87.48 | 89.47 | 85.28 |
| K2 words (1001-2000) | Families | 30 | 22 | 25 | 20 |
| | Types | 35 | 23 | 28 | 21 |
| | Tokens | 57 | 33 | 45 | 36 |
| | % | 6.4 | 3.97 | 3.98 | 3.73 |
| AWL words (Academic Word List) | Families | 17 | 14 | 16 | 19 |
| | Types | 22 | 15 | 20 | 22 |
| | Tokens | 33 | 22 | 35 | 40 |
| | % | 3.7 | 2.65 | 3.1 | 4.15 |
| Off-list words | Families | N/A | N/A | N/A | N/A |
| | Types | 34 | 30 | 25 | 44 |
| | Tokens | 45 | 49 | 39 | 66 |
| | % | 5.05 | 5.9 | 3.45 | 6.84 |

Note: In the output text, punctuation is eliminated; all figures are replaced by the word number; contractions are replaced by constituent words; in the 1k sub-analysis content + function words may sum to less than total (depending on treatment of proper nouns and decision to class numbers as 1k although not contained in 1k list); single letters are eliminated except for 'a' and 'I.'

As the table shows, test-takers 4 and 6 produced the most off-list words, such as proper nouns, abbreviations, and acronyms. It is worthwhile to note that these two test-takers were the only ones out of the four test-takers who received lower scores in the assessment of lexical range in the face-to-face speaking tests as compared to their scores assigned in the audio-recorded speaking tests (0.25 and 0.5, respectively).

4.3 Thematic Analysis and Frequency of Nonverbal Data

To examine the data collected from the open-ended questions in the questionnaire and the video-stimulated verbal recall sessions, an analysis was conducted.

The recall sessions were semi-structured; i.e., they were conducted following a previously designed script that allowed for follow-up questions based on the thoughts shared by examiners. Before reviewing the video recording of the speaking tests they had conducted, I asked examiners if they could remember what their thoughts were prior to the start of the test. While examiners A and F highlighted that their focus was on the structure of the test and the questions they would be asking, examiner D stated that the main concern was with possible changes in behaviour while conducting a standardized test (such as being less natural), and examiner E shared thoughts about initial impressions of test-taker 5 (“warm” and “open”).

Upon conclusion of all speaking tests and the transcription of the data, I conducted a thematic analysis (Braun & Clarke, 2006) to identify patterns and, later, establish themes. The results and discussion presented in the next section include the themes identified in the data and the results of the analysis of the nonverbal cues produced during the face-to-face speaking tests.

4.3.1 Perception of test-takers’ overall behaviour

During the video-stimulated verbal recall sessions, to garner information about how examiners perceived test-takers, I asked examiners to share their thoughts regarding the test-takers’ overall behaviour, smiles, gestures, and eye contact when the face-to-face speaking tests took place.

Examiners’ thoughts presented a range of different opinions about test-taker’s behaviour. For example, in Excerpt 1, examiner E discussed the seemingly natural behaviour of test-taker 5 during the test.

Excerpt 1

“Her behaviour seems really natural to me for a student who’s in a slightly uncomfortable situation. She was polite, and cheerful, and smiling, and trying to do the best she could in a somewhat strange situation. [. . .] It’s not natural. it’s not comfortable, but she had that really good, um... hand movements, head movement, her face, her smile.”

The examiner’s account showed not only that the general behaviour of the test-taker was salient in the constraints of the exam setting but also that the examiner decoded and evaluated it, in this case perceiving it as positive. The examiner did not elaborate on the definitions of “natural” or “really good hand movements”. As excerpt 2 shows, examiner D also recalled thoughts of test-taker 4’s behaviour.

Excerpt 2

“I sensed, and I don’t know if it was... what in her body language that she was a little more unsure in this part.”

As the excerpt shows, examiner D decoded the changes in test-taker 4’s behaviour as an indication of changes in the test-taker’s level of confidence during part 3 of the speaking tests. Examiner D’s comments also indicated that the examiner perceived changes in the test-taker’s behaviour throughout the test. As Excerpt 3 illustrates, in the examiner’s opinion, changes seemed to be related to whether the test-taker’s answers were more generic or personal.

Excerpt 3

“I think she was most comfortable... she’s actually really interestingly... I notice a difference; I felt she was more comfortable when she was talking about her hometown and the sort of... she... it seems nostalgic almost.”

The examiners' accounts of how they perceived the test-takers' behaviour revealed that they were not impervious to the test-takers' overall behaviour during the face-to-face speaking tests, an observation that is supported by the fact that the examiners spontaneously shared their thoughts and considerations about the test-takers' behaviour. It should be noted, however, that the examiners neither provided any further information on what in the observable behaviour of test-takers led to the perception of the test-takers as being "comfortable" or nor discussed how the interplay between attitude, perception, and behaviour could have led to the assessment of their behaviour as "natural."

4.3.2 Frequency of production of test-takers' nonverbal cues

Having identified that examiners were conscious of the test-takers' nonverbal cues, I conducted an analysis of the three nonverbal cues that are the focus of this study. Results showed that test-takers produced a total of 1,112 observable cues during the face-to-face speaking tests, with gestures being the most frequent (48.74%). To contextualize the thoughts the examiners had shared about the test-takers, I analyzed the frequency of nonverbal cues generated by each test-taker against the total number of nonverbal cues produced by all test-takers. Results showed a wide range in the frequency of generated cues.

Table 14

Frequency of Observable Nonverbal Cues per Test-taker

| Test-takers | Smiles | % | Gestures | % | Eye contact | % | Totals | % |
|--------------|--------|-------|----------|-------|-------------|-------|--------|-------|
| Test-taker 1 | 30 | 20.13 | 77 | 14.21 | 110 | 26.13 | 217 | 19.51 |
| Test-taker 4 | 29 | 19.46 | 71 | 13.10 | 86 | 20.43 | 186 | 16.73 |
| Test-taker 5 | 34 | 22.82 | 136 | 25.09 | 97 | 23.04 | 267 | 24.01 |
| Test-taker 6 | 56 | 37.58 | 258 | 47.60 | 128 | 30.40 | 442 | 39.75 |
| Totals | 149 | | 542 | | 421 | | 1112 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

According to the results in Table 14, test-takers produced between 186 to 442 nonverbal cues. Test-taker 6's frequency of observable cues (39.75%) was more than twice as high as that of test-taker 4 (16.73%) and is the highest overall. The number of gestures generated by this participant during the face-to-face speaking test (47.60%) was also more than three times higher than those produced by test-takers 1 (14.21%) and 4 (13.10%).

Next, to learn which cues each test-taker produced more/less frequently, I calculated the frequency of each observable nonverbal cue generated by each test taker in the context of the total production per each test-taker (see Table 15).

Table 15

Frequency of Observable Nonverbal Cues per Test-taker (in Proportion to the Total Cues Produced per Test-taker)

| Nonverbal cue | Test-taker 1 | % | Test-taker 4 | % | Test-taker 5 | % | Test-taker 6 | % | Totals | % |
|---------------|--------------|-------|--------------|-------|--------------|-------|--------------|-------|--------|-------|
| Smiles | 30 | 13.82 | 29 | 15.59 | 34 | 12.73 | 56 | 12.67 | 149 | 13.40 |
| Gestures | 77 | 35.48 | 71 | 38.17 | 136 | 50.94 | 258 | 58.37 | 542 | 48.74 |
| Eye contact | 110 | 50.69 | 86 | 46.24 | 97 | 36.33 | 128 | 28.96 | 421 | 37.86 |
| Totals | 217 | | 186 | | 267 | | 442 | | 1112 | |

Results also showed that test-takers 1 and 4 generated mostly eye contact (50.69% and 46.24%, respectively) in proportion to the number of nonverbal cues they generated during the face-to-face speaking tests while test-takers 5 and 6 produced gestures more than any other cue (50.94% and 58.37%, respectively). Frequency of production of smiles was the lowest of all cues for all test-takers, ranging from 12.67% (test-taker 6) to 15.59% (test-taker 4).

4.3.3 Perception of test-takers' eye contact.

In the recall sessions, examiners shared their thoughts about how they perceived test-takers' eye contact during the face-to-face speaking tests. Examiners' accounts showed their awareness of test-takers' eye contact throughout the test. For instance, in the excerpt below, examiner D stated feeling impressed with what was perceived as prolonged eye contact.

Excerpt 4

“I thought she was looking at me directly most of the time. I was actually impressed with her eye contact.”

Examiner D's report also indicated that the examiner associated the test-taker's change in eye contact with changes in the test-taker's confidence level as Excerpt 5 illustrates.

Excerpt 5

“Uh, often when students are nervous, they don't make eye contact [. . .] So it just showed confidence to me.”

“Like there was something different in her eye contact in this part of the test that she seemed a little less confident.”

Examiner E's thoughts shared during the recall session showed that the examiner perceived the test-taker's eye contact during part 1 of the test as natural, which was identified as being “logical, Canadian style, or test-taking” style.

Excerpt 6

“In the first section, I feel her eye contact with me was natural and logical and not noticeable in any way. It was a Canadian style, or it was a test-taking style; there wasn’t anything special.”

Examiner E’s account, which suggests that the test-taker’s eye contact was not particularly special or unusual, is relevant as, coming from different cultural backgrounds and being in a zero-acquaintance situation, the perception of the eye contact as “Canadian” might have led to a more positive perception of the test-taker.

Excerpt 7 illustrates how examiner E perceived the changes in the test-taker’s eye contact as a sign of how focused the test-taker was during the second part of the test. The fact that the test-taker did not look towards the examiner during that stage of the test was not, therefore, considered negative but rather a natural reaction due to the context in which the interaction was taking place.

Excerpt 7

“It was in the second part where I noticed she was really focused. [. . .] And she looked down for almost all of her air-pollution 2 minutes, which is totally fine. It didn’t bother me at all, but I noticed it whereas in this section I didn’t notice any eye contact.”

Examiner E’s further explanation described how the belief that the level of difficulty of part two might have resulted in test-taker 5 looking away more frequently. As Excerpt 8 illustrates, the apparent difficulty the test-taker might have had with the task was perceived by the examiner as being the result of a lack of personal connection between the test-taker and the topic.

Excerpt 8

Researcher: “You said you noticed a change in her behaviour in this second part. Do you remember what you thought or if you noticed it then?”

Examiner E: “I noticed it then because I thought, in the first section, our eye contact was almost constant. [. . .] In this part, I feel like, perhaps, it was more difficult for her because it wasn’t so personal. It was more of a topic out there that she felt at the beginning...looking down helped her remember what she was supposed to say, and then, when she was eventually able to look up and see me, she realized, ‘OK, I’m on the right track; I can continue. I don’t need to look down the whole time.’”

Examiner F’s thoughts about test-taker 6’s eye contact differed from those shared by the other examiners. In the recall session, the comments this examiner made indicated that the examiner perceived the test-taker’s eye contact as a tool to communicate nonverbally during the speaking test.

Excerpt 9

“She looked at me with this perplexed look just for a second as though ‘are you coming into this conversation or not?’ It was as though she was like, ‘are you going to comment?’”

This examiner shared a similar thought later in the recall session as illustrated in Excerpt 10.

Excerpt 10

“I just sort of decided... ‘Okay, I am just going to listen and see how far she goes with this.’ But there’s a moment you can see on her face where she pauses for a second and gives me a look like, ‘are you going to say anything?’”

4.3.4 Frequency of production of test-takers’ eye contact.

The examiners’ accounts suggested that not only were they aware of the test-takers’ eye contact, but that they also interpreted them during the face-to-face exchanges. An analysis of the

observed eye contact showed that test-takers looked²⁰ at examiners 421 times. Test-takers 1 and 6 generated the highest frequency of eye contact (26.13% and 30.40%, respectively).

Next, I analyzed the frequency of production of observable eye contact alongside speech or in isolation. Results from the analysis of the cues generated in isolation indicated minimal differences in the frequency of production of all test-takers, with frequencies ranging between 23.08% and 26.92%.

Table 16

Frequency of Observable Mutual Eye Contact in Isolation/alongside Speech

| Test-takers | Eye contact in isolation | % | Eye contact alongside speech | % | Total | % |
|--------------|-----------------------------|-------|---------------------------------|-------|-------|-------|
| Test-taker 1 | 18 | 23.08 | 92 | 26.82 | 110 | 26.13 |
| Test-taker 4 | 21 | 26.92 | 65 | 18.95 | 86 | 20.43 |
| Test-taker 5 | 20 | 25.64 | 77 | 22.45 | 97 | 23.04 |
| Test-taker 6 | 19 | 24.36 | 109 | 31.78 | 128 | 30.40 |
| Totals | 78 | | 343 | | 421 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

The analysis of the frequency of mutual eye contact alongside speech, however, indicated differences in the frequency of eye contact produced by test-takers. Results showed that test-takers 1 and 6, who received the lowest overall scores in the face-to-face speaking tests, generated the highest instances of eye contact (26.82% and 31.78%, respectively).

4.3.5 Perception of test-takers' gestures

During the recall sessions, examiners reported on the test-takers' use of hand gestures

²⁰ The frequency count indicates the number of times examiners and test-takers looked at each other.

during the face-to-face speaking tests. Their accounts did not refer to gestures by type but rather to observations about the use of hand gestures in general.

Accounts from the examiners showed that each examiner perceived gestures produced by examiners. For instance, as Excerpt 11 illustrates, examiner E's referred to the test-taker's confidence level and use of gestures, which were considered "good." The examiner, however, did not expand on what constitutes "good" or how the gestures were evaluated as such.

Excerpt 11

"Her behaviour seems really natural to me for a student who's in a slightly uncomfortable situation. [. . .] but she had that really good, um... hand movements, head movement, the face, her smile."

Similar to the account from examiner E, examiner F also reported interpreting the gestures produced by test-taker 6 during the speaking test. For instance, during task 2, examiner F felt that test-taker 6 possibly had extensive knowledge about the topic being discussed. Excerpt 12 illustrates how the gestures generated during the test might have influenced how the examiner perceived the test-taker's overall knowledge about the topic.

Excerpt 12

"I can also see her hand gestures are as though she's lecturing me. And in the first one she appeared more nervous; her hand gestures were more nervous. And in this it's as though she's got sort of a structured idea in her mind and she's just going through the structure, but she's telling me. So she's using her hands like this to tell me about something she's informed about."

Examiner F shared a similar thought later in the recall session, as illustrated in Excerpt 13.

Excerpt 13

"Um, well, she's got those same hand gestures again where she starts... she looks as

though she's an authority on the topic.”

The gestures produced by test-taker 6 were also seen as a tool to help trigger the test-taker's memory. In this case, however, the uniqueness of the gestures seemed to distract the examiner during the test (Excerpt 14).

Excerpt 14

“So I guess that helps he trigger, but I didn't... [. . .] I just thought it was... I've never seen somebody do that before. I've seen people do other things when they're looking for a particular word. They probably don't even know they do, but at that point in time, yeah... I sort of was somewhat distracted by that...”

Examiner F's reports showed awareness of the gestures generated by test-taker 6, which could also indicate their possible influence on how this examiner perceived the test-taker. This possible influence was supported by one of the examiner's last accounts during the recall session.

Excerpt 15

Examiner F: “I think because I have a limited capacity for concentration, so I was trying to focus specifically. It's a language test... I was trying to focus on her language rather than... when I do this with my students at the school, I rarely look up. [. . .] I look up occasionally to see where their eyes are. [. . .] Instead, what I do is I have my head down and I'm taking notes [. . .] So I tend to not look at them, and I think that allows me to focus more on their spoken language.”

4.3.6 Frequency of production of test-takers' gestures

To contextualize the examiners' accounts, I conducted an analysis of the observable gestures produced by the test-takers during the speaking tests. As previously discussed, results (Table 13) showed that test-taker 6 generated 47.60% of the observable gestures while test-taker

4 generated the least number of gestures (13.10%). Although results revealed differences among the test-takers, questions remained about the types of gestures that the four test takers generated. Knowledge from the analysis of the subcategories (*self- and object-adaptors* and *illustrators*) could allow for a better understanding of their role in the assessment of the test-takers' performances during the speaking tests. I therefore divided the category into the subcategories *illustrators* and *self- and object-adaptors*. Results showed that test-takers generated 542 observable gestures, and over 70% of those were illustrators (see Table 17). Results also showed a small difference in the number of self- and object-adaptors generated by the test-takers.

Table 17

Frequency of Overall Observable Gestures by Type

| Gestures | Gestures | % |
|-----------------|----------|-------|
| Illustrators | 383 | 70.66 |
| Self-adaptors | 74 | 13.65 |
| Object-adaptors | 85 | 15.68 |
| Total | 542 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

To gain a deeper understanding of the examiners' perceptions of each test-taker's gestures, I analyzed the frequency of production per test-taker. Unlike the holistic analysis of the subcategories, the individual analysis showed differences in the frequency of gestures produced by test-takers (see Table 18).

Table 18

Frequency of Observable Gestures by Type per Test-taker

| Test-takers | Illustrators | % | Self-adaptors | % | Object-adaptors | % | Totals | % |
|--------------|--------------|-------|---------------|-------|-----------------|-------|--------|-------|
| Test-taker 1 | 25 | 6.53 | 23 | 31.08 | 29 | 34.12 | 77 | 14.21 |
| Test-taker 4 | 59 | 15.40 | - | - | 12 | 14.12 | 71 | 13.10 |
| Test-taker 5 | 118 | 30.81 | 16 | 21.62 | 2 | 2.35 | 136 | 25.09 |
| Test-taker 6 | 181 | 47.26 | 35 | 47.30 | 42 | 49.41 | 258 | 47.60 |
| Totals | 383 | | 74 | | 85 | | 542 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

Table 18 reveals that test-taker 6 generated 47.26% of all illustrators, 47.30% of self-adaptors, and 49.41% of object adaptors. Test-taker 5 produced the second highest number of illustrators (30.81%), and test-taker 1 the highest number of self- and object-adaptors (31.08% and 34.12%, respectively). Test-taker 4 did not produce any observable self-adaptors.

Next, an analysis of the gestures that were generated alongside speech and in isolation was conducted. Results showed that, out of the 542 observable gestures, only 28 were produced in isolation. Test-taker 5 generated the highest frequency of self-adaptors (71.43%) and test-takers 1 and 6 the most object-adaptors (35.71% each) (see Table 19).

Table 19

Frequency of Gestures in Isolation (I/I)

| Test-takers | Illustrators | | Self-adaptors | | Object-adaptors | | Totals | % |
|--------------|--------------|---|---------------|-------|-----------------|-------|--------|-------|
| | I/I | % | I/I | % | I/I | % | | |
| Test-taker 1 | 0 | - | 2 | 14.29 | 5 | 35.71 | 7 | 25.00 |
| Test-taker 4 | 0 | - | 0 | - | 3 | 21.43 | 3 | 10.71 |
| Test-taker 5 | 0 | - | 10 | 71.43 | 1 | 7.14 | 11 | 39.29 |
| Test-taker 6 | 0 | - | 2 | 14.29 | 5 | 35.71 | 7 | 25.00 |
| Totals | 0 | | 14 | | 14 | | 28 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

The frequency of generated gestures alongside speech showed that test-taker 6 produced the most illustrators (47.26%), self-adaptors (55.0%), and object-adaptors (52.11%). Results also show that test-taker 1 generated the second most frequent number of self-adaptors (35.0%) and object-adaptors (33.80%) while test-taker 4 did not generate any observable self-adaptor (see Table 20).

Table 20

Frequency of Gestures alongside Speech (A/S)

| Test-takers | Illustrators | | Self-adaptors | | Object-adaptors | | Totals | % |
|--------------|--------------|-------|---------------|------|-----------------|-------|--------|-------|
| | A/S | % | A/S | % | A/S | % | | |
| Test-taker 1 | 25 | 6.53 | 21 | 35.0 | 24 | 33.80 | 70 | 13.62 |
| Test-taker 4 | 59 | 15.40 | 0 | - | 9 | 12.68 | 68 | 13.23 |
| Test-taker 5 | 118 | 30.81 | 6 | 10.0 | 1 | 1.41 | 125 | 24.32 |
| Test-taker 6 | 181 | 47.26 | 33 | 55.0 | 37 | 52.11 | 251 | 48.83 |
| Totals | 383 | | 60 | | 71 | | 514 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

4.3.7 Perception of test-takers' smiles.

During the video-stimulated recall sessions, examiners spontaneously shared thoughts and perceptions they recalled about the test-takers' smiles. Prior to the beginning of the first speaking test, examiner A mentioned test-taker 1's "beautiful and warm smiles." I noted this comment in one of the notes taken during the data collection stage and used it (as well as the other notes) to inform the coding of the data. Although this examiner did not share any further thoughts after this initial comment, the remark showed the examiner's awareness of and reaction to the test-taker's smiles. Examiners' reports suggest possible awareness of the differences in smiles produced by the test-takers. In the excerpt below, for example, examiner D commented on the type of smile generated by test-taker 4 and the possible connection with the test-taker's intention to make a joke.

Excerpt 16

"I think the other sort of smile was that she talked about something personal instead of something abstract. So she started, 'ohh, it's my birthday', and she kind of made a joke, right?"

In the excerpt above, examiner D also mentioned how she perceived differences in smiles depending on whether test-taker 4 was talking about a general or a more personal topic. This examiner made similar observations throughout the recall session.

Examiner E also linked smiles to a more natural behaviour. The examiner shared thoughts about test-taker 5's smiles, which were considered natural despite the fact that the testing context usually leads test-takers to feeling uncomfortable. Excerpt 17 illustrates the examiner's perceptions.

Excerpt 17

“Her behaviour seems really natural to me for a student who’s in a slightly uncomfortable situation. She was polite and cheerful and smiling, and trying to do the best show could in a somewhat strange situation [. . .] It’s not natural, and it’s not comfortable, but she had really good, um. . . hand movements, head movement, the face, her smile.”

Examiner E’s detailed account gave an insight into how aware this examiner was of the nonverbal cues the test-taker generated during the face-to-face speaking test and of the examiner’s assessment of these cues during the speaking test.

Examiner F also shared thoughts about test-taker 6’s smiles, which, at times, the examiner identified as turn-taking cues. In the excerpt below, when asked about the test-taker’s smile, the examiner mentioned thinking that the test-taker was using smiles to indicate that she had finished answering a question.

Excerpt 18

“I remember thinking, ‘okay, she’s giving me a cue - I’m done with that question’. I was fairly impressed, but I couldn’t say that I was impressed. It felt unnatural for me not to say ‘wow’ or to ask some more natural follow-up questions.”

Examiner F’s account also indicated the possibility that the use of smiles by this test-taker could have influenced how the examiner assessed test-taker 6’s performance at the time of the speaking test. The use of words such as “impressive” and “wow” signal a conscious assessment of the test-taker’s use of that cue as a tool in communicative events. Test-takers’ smiles were, therefore, mentioned by all examiners and, in at least one instance, was perceived positively during the exchanges that took place in the face-to-face speaking tests.

4.3.8 Frequency of production of test-takers' smiles

The analysis of the smiles generated during the speaking tests shows that these cues were produced 149 times. Test-taker 6 generated the highest number of smiles (37.58%) and test-taker 4 the least (19.46%) (Table 13). To examine the influence of different smiles in the assessment of test-takers' performances, the category was broken down into the subcategories *Duchenne* and *non-Duchenne* smiles. Results showed that non-Duchenne smiles comprised 54.36% and Duchenne smiles 45.64%.

The next step of the analysis consisted of the examination of the frequency of Duchenne and non-Duchenne smiles per test-taker. Results revealed that test-taker 6 produced the most Duchenne smiles (32.35%) and test-taker 1 produced the least (11.76%). On the other hand, test-takers 1 and 6 produced 27.16% and 41.96% of all non-Duchenne smiles, respectively, the two highest frequencies in this subcategory (see Table 21).

Table 21

Frequency of Production of Smiles by Type

| Test-takers | Duchenne | | Non-Duchenne | | Totals | % |
|--------------|----------|-------|--------------|-------|--------|-------|
| | smiles | % | smiles | % | | |
| Test-taker 1 | 8 | 11.76 | 22 | 27.16 | 30 | 20.13 |
| Test-taker 4 | 17 | 25.0 | 12 | 14.81 | 29 | 19.46 |
| Test-taker 5 | 21 | 30.88 | 13 | 16.05 | 34 | 22.82 |
| Test-taker 6 | 22 | 32.35 | 34 | 41.96 | 56 | 37.58 |
| Totals | 68 | | 81 | | 149 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

Following the analysis of each subcategory, I examined the smiles produced alongside speech or in isolation. Results revealed that test-takers 4 and 5 produced the most Duchenne (38.46% and 30.77%, respectively) and non-Duchenne smiles (31.58% each) in isolation (see

Table 22).

Table 22

Frequency of Observable Duchenne and non-Duchenne Produced in Isolation (I/I)

| Test-takers | Duchenne I/I | | Non-Duchenne | | Totals | % |
|--------------|--------------|-------|--------------|-------|--------|-------|
| | I/I | % | I/I | % | | |
| Test-taker 1 | 3 | 11.54 | 3 | 15.79 | 6 | 13.33 |
| Test-taker 4 | 10 | 38.46 | 6 | 31.58 | 16 | 35.56 |
| Test-taker 5 | 8 | 30.77 | 6 | 31.58 | 14 | 31.11 |
| Test-taker 6 | 5 | 19.23 | 4 | 21.05 | 9 | 20.0 |
| Totals | 26 | | 19 | | 45 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

In terms of smiles produced alongside speech, results showed that test-takers 5 and 6 produced the most Duchenne smiles (30.85% and 40.48%, respectively). Contrastively, test-takers 1 and 6 produced, altogether, 79.04% of the non-Duchenne smiles generated alongside speech. Results can be found in Table 23. Results showed that test-takers 5 and 6 produced the most Duchenne smiles (30.85% and 40.48%) and test-takers 1 and 6 the most non-Duchenne smiles (30.65% and 48.39%, respectively).

Table 23

Frequency of Observable Duchenne and Non-Duchenne Smiles Produced alongside Speech (A/S)

| Test-takers | Duchenne smiles A/S | % | Non-Duchenne smiles A/S | % | Totals | % |
|--------------|------------------------|-------|----------------------------|-------|--------|-------|
| Test-taker 1 | 5 | 11.90 | 19 | 30.65 | 24 | 23.08 |
| Test-taker 4 | 7 | 16.67 | 6 | 9.68 | 13 | 12.50 |
| Test-taker 5 | 13 | 30.95 | 7 | 11.29 | 20 | 19.23 |
| Test-taker 6 | 17 | 40.48 | 30 | 48.39 | 47 | 45.19 |
| Totals | 42 | | 62 | | 104 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

4.3.9 Encouragement through nonverbal cues.

During the recall sessions, examiners shared accounts about the use of nonverbal cues to encourage test-takers in speaking-test situations. Reports revealed that nonverbal cues were used to communicate with the test-takers while avoiding the use of any verbal input. One examiner observed, after the conclusion of the recall session, that nonverbal communication was used to make test-takers more comfortable during the IELTS Speaking Test as the guidelines established by IELTS prevents examiners from veering away from the test script. For instance, Excerpt 19 illustrates how examiner A described the use of nonverbal cues during the speaking test.

Excerpt 19

“I was nodding to encourage her because in an IELTS exam, the examiner isn’t allowed to say anything. So it was just encouragement rather than anything else, I think.”

Examiner A also described using smiles to help the test-taker feel more comfortable.

Excerpt 20

Researcher: “I noticed you were smiling more to her and back to you at that point, and

um, do you remember why?"

Examiner A: "Cause you can't say anything in an IELTS test. You can only follow the script, so encouragement has to be nonverbal. And I think if you smile at somebody or nod, you know... it gives cues that you're listening and paying attention and hopefully they'll feel better."

Examiner D provided similar accounts when describing the use of smiles and nodding in speaking test contexts.

Excerpt 21

"I think it is my...my involuntary sort of 'you can do it; it's okay, come on now. Yeah."

Researcher: "You're smiling." (in the video)

Examiner D: "[. . .] it's my defence when she makes a mistake [. . .] And so my default when students make mistakes...because I find students have a really hard time making mistakes 'cause they get panicked."

Examiner E shared thoughts on the use of nonverbal cues to encourage test-takers. For this examiner, nodding is used to support students during speaking tests.

Excerpt 22

"And also, when I nod, it, to me, signals to the student 'you're doing well, keep going, you're doing well, keep going... just keep going, you're doing well."

Examiner F, who shared concerns about focusing on the testing procedures, also used nonverbal cues to encourage test-taker 6. During the recall session, the examiner shared that, at the beginning of the test, a mistake made by the test-taker resulted in the production of smiles and nodding to try to put her at ease (Excerpt 23).

Excerpt 23

Researcher: “I noticed that, at that time, you were smiling at her and nodding at her.”

Examiner F: “I try to make people feel as comfortable as possible, and so I remember knowing she was feeling nervous. And, especially, she fumbled right away with me asking her name, so I just wanted to make her feel more comfortable. I can’t help it.”

Examiner F’s thoughts also provided an insight into how the examiners’ understanding of the expected behaviour of certified IELTS examiners. For example, as Excerpt 24 illustrates, examiner F stated how the use of smiles might not be acceptable in an official testing situation.

Excerpt 24

“When I’m talking with somebody, it’s really hard not to give them those nonverbal cues... just to communicate with somebody, and I think just a smile allows them to talk more and encourages them to talk more, so I have a hard time stifling that, and I knew I wasn’t supposed to.”

Similarly to the account of examiner F, examiner A also shared thoughts and opinions about the use of nonverbal cues in official IELTS Speaking Tests.

Excerpt 25

“I was nodding to encourage her because in an IELTS exam the examiner isn’t allowed to say anything. So, it was just encouragement rather than anything else, I think. Yeah.”

While describing the use of nonverbal cues to encourage test-takers in speaking test situations, examiner A shared some thoughts on how mimicking an interlocutor’s nonverbal cues could be an unconscious attempt to try to help test-takers relax.

Excerpt 26

“Um, yeah, I think that’s kind of reflecting people’s body language. I think you do it unconsciously, and it shows that you’re trying to put them at ease, right?”

Mimicking, the reproduction of an individual's nonverbal behaviour, was mentioned as an element that could influence how examiners interacted with test-takers. To examine if the frequency of mimicked cues could have played a role on the examiners' assessment of test-takers' performances in the face-to-face speaking tests, an analysis of mimicked cues was conducted. Results of the analysis can be found in Table 24.

Table 24

Nonverbal Cues Mimicked by Examiners

| Participants | Observable nonverbal cues generated by test-taker | Cues mimicked by examiner | % |
|--------------|---|---------------------------|------|
| Examiner A | 217 | 13 | 5.99 |
| Examiner D | 186 | 17 | 9.14 |
| Examiner E | 267 | 17 | 6.37 |
| Examiner F | 442 | 6 | 1.36 |
| Totals | 1112 | 53 | |

Note. Percentage value in proportion to the total number of cues produced by all test-takers.

As the results showed, examiner F's frequency of mimicry of cues as the lowest (1.36%) and examiner D's the highest (9.14%).

Results from the analysis of the data indicated that examiners were aware of some of the nonverbal cues generated in the face-to-face speaking tests; those cues were interpreted within the context of the speaking test and, based on the examiners' accounts, may have resulted in possible judgments regarding the test-takers knowledge of the topic and comfort and confidence levels. Results from the analysis of nonverbal cues vis-à-vis the scores assigned also suggested possible connections between the variation in the frequency and types of generated cues and the differences in the scores assigned to the face-to-face and audio-recorded performances.

Chapter 5: Discussion

The first part of this chapter presents a discussion of the key findings followed by the implications and limitations of this study. Lastly, future research directions and a brief conclusion are presented.

5.1 Key Findings

This study aimed to investigate the role of smiles, gestures, and eye contact on the scores assigned to test-takers by EL1 examiners during simulated IELTS Speaking Tests. Previous research on the assessment of test-takers' speaking skills has shown that variables such as examiners' exposure to foreign languages, gender, and cultural similarities between examiners and test-takers may affect how examiners assess test-takers' performances (Briegel-Jones, 2014; O'Loughlin, 2000; O'Sullivan & Lu, 2002). The present study was motivated by a gap in previous research, which, to the best of my knowledge, has yet to investigate the role of nonverbal cues produced by test-takers in the scores assigned during simulated IELTS Speaking Tests. Results from the analysis of the data indicated differences in the scores assigned to the test-takers' performances during the face-to-face and audio-recorded speaking tests.

5.1.1 Observable nonverbal cues vis-à-vis test-takers' performances.

To address the research question, qualitative and quantitative analyses were conducted. Overall, test-takers produced 1,112 smiles, gestures, and eye contact during the four face-to-face speaking tests.

Eye Contact

Findings from the analysis of test-takers' eye contact show that mutual eye contact varied in frequency between 20.43% (test-taker 4) and 30.40% (test-taker 6). Findings from the analysis

of frequency of production in isolation show a small difference among all test-takers (between 23.08% and 26.92%), with test-takers 1 and 6 generating the lowest frequency (23.08% and 24.36%, respectively). Results from the analysis of the cues alongside speech, however, painted a different picture. While speaking, test-takers 1 and 6, who received the lowest scores for their performances in the face-to-face speaking test, generated 26.13% and 30.40% of all instances of eye contact, the highest frequency of all test-takers. These findings are supported by previous studies on eye contact/face gazing and perception of speakers. Longer periods of eye contact (i.e., lower frequency of production) have been linked to speakers being perceived as more intelligent and more competent in North America (Imada & Hakel, 1977; Sodikoff, Firestone, & Kaplan, 1974; Wheeler et al., 1979). Eye contact has also been shown to help establish interpersonal relationships in social interaction (Csibra & Gergely, 2006; Kleinke, 1986; Senju & Johnson, 2009).

Cultural disparities in regards to the frequency of eye contact were not mentioned in the recall sessions; however, review of the data indicates that some interviewers seemed surprised with test-takers' eye contact within the speaking test context. However, in the case of test-taker 5 (who received the highest scores in the face-to-face speaking tests), examiner E shared perceiving the test-taker's eye contact as "Canadian" style, which was considered "natural." It is possible to postulate that perceived similarities between the eye contact produced by the test-taker and those produced by the examiner could have influenced the assessment of the test-taker's performance. This argument is supported by previous research on second language assessment, which has shown that speakers that generate nonverbal cues that are perceived as appropriate or familiar by native speakers seem to be able to compensate for possible linguistic weaknesses (Jenkins & Parra, 2003).

Gestures

Results from the analysis of observable gestures showed that test-takers produced 542 hand gestures, with test-taker 6 generating 47.60% of those cues. An examination of the subcategories revealed that test-takers 5 and 6 used the most illustrators (30.81% and 47.26%, respectively). A comparative analysis between the frequency of production of this subcategory and the scores assigned showed that the two test-takers who received the lowest scores on the face-to-face speaking tests generated both the most and the least number of illustrators (test-taker 1, 6.53% and test-taker 6, 47.26%). It is possible to postulate that the scarcity of production of illustrators by test-taker 1 could have resulted in this test-taker receiving the second lowest score of all face-to-face speaking tests. Previous research has suggested that gestures may enhance comprehension and may be generated with the intention of communicating with an interlocutor (Bavelas et al., 2002; Kelly et al., 1999). Furthermore, illustrators have been found to improve interlocutors' attention and help clarify speakers' discourse across cultures (Maricchiolo et al., 2009). This argument, however, could not be applied to the assessment of test-taker 6's performance. Results from the analysis of the qualitative data helped shed some light on the possible reasons for the unexpected finding. As previously mentioned, examiner F disclosed (Excerpt 15), during the recall session, a preference for avoidance of visual input during the assessment of students' speaking skills as that was considered a distraction. As per the examiner's account, this perceived interference on the proper assessment of students' speaking skills has resulted in the examiner's choice to not look at students in an attempt to properly evaluate their speaking skills. This preference might provide some insight into the evaluation of test-taker 6's performance, especially when it is taken into account that the examiner shared that one of the illustrator gestures used by the test-taker was quite "distracting" due to its uniqueness. Based on these findings, it is possible to postulate that examiner F's preference for no visual input might have

influenced how the examiner assessed test-taker 6's performance in the face-to-face speaking test and might explain why that examiner assigned higher scores to the test-taker's audio-recorded performance. Another possible argument is that the high frequency of illustrators generated by the test-taker could have been perceived by the examiner as an attempt to compensate for limited lexical resource. This argument is supported by previous research that investigated the compensatory function of gestures; results indicated that speakers increase the use of gestures when they lack vocabulary to provide accurate descriptions (Bavelas et al., 2002; Gullberg, 2006). Finally, it is important to take into consideration that differences in the cultural backgrounds of the test-taker and the examiner could have resulted in difficulty in the decoding of the gestures, and, therefore, possibly influenced the evaluation of her performance in the face-to-face speaking test. Such a possibility should not be overlooked if we consider that the examiner stated not having seen gestures like the ones used by the test-taker before. Despite the automaticity of the production of illustrators, research has shown that gestural form and use conventions are influenced by cultural norms (Gullberg, 2006). Previous research has also shown that different gestures may be decoded differently by interlocutors from different cultural backgrounds (Jokinen & Allwood, 2010; Knapp & Hall, 2006). Therefore, it is possible, as examiner F stated, that the unique illustrators used by test-taker 6 might have caused distractions during the assessment of her performance in the face-to-face speaking test.

Although examiners did not share specific thoughts or observations regarding the adaptors generated during the face-to-face tests, results showed that test-takers produced 74 self-adaptors and 85 object-adaptors, with test-takers 1 and 6 producing the highest frequency of those cues (31.08% and 47.30% self-adaptors and 34.12% and 49.41% object-adaptors, respectively). A comparative analysis of these findings and the scores assigned showed that test-takers 1 and 6 also received the lowest scores in the face-to-face speaking tests. Previous studies corroborate

these findings. Adaptors are unintentional gestures that are perceived as being linked to negative feelings and the perception of speakers as anxious, nervous, self-conscious, and even deceptive (Henningsen, Valde, & Davies, 2005; Richmond & McCroskey, 2000). Studies have also suggested that interlocutors may use object- and self-adaptors to make dispositional inferences as it is believed that adaptors may reflect unconscious feelings and thoughts (Mahl, 1968; Ekman & Friesen, 1974). In the context of second language speaking, research has shown that anxious EAL speakers increase the use of self- and object-adaptors (Gregersen, 2005). Therefore, it is possible to postulate that the high number of adaptors produced by test-takers 1 and 6 may have influenced how examiners A and F assessed the test-takers' performances in the face-to-face speaking tests.

Smiles

Test-takers generated a total of 68 observable Duchenne and 81 non-Duchenne smiles throughout the four speaking tests. A comparative analysis of the scores assigned and the observable smiles showed that the two test-takers who received the lowest scores in the face-to-face speaking tests (test-takers 1 and 6) generated the most non-Duchenne smiles overall. Test-taker 6 also produced the highest number of Duchenne smiles of all test-takers; however, the frequency of non-Duchenne smiles surpassed that of Duchenne; non-Duchenne smiles comprised more than 60% of the smiles produced by that test-taker. Findings from the analysis of the qualitative data suggested that examiners were not impervious to the differences in smiles being produced during the speaking test, commenting on test-takers' Duchenne smiles, which were seen as a sign of confidence and engagement. Questions remained, however, regarding test-taker 6, who received lower scores during the face-to-face speaking test when compared to the scores assigned to her audio-recorded performance. Research has shown that interlocutors from North American and European countries perceive individuals who generate Duchenne smiles as more

intelligent, likeable, and trustworthy when compared to individuals who generate non-Duchenne smiles (Gunnery & Ruben 2016; Johnston, Miles, & Macrae, 2010; Krumhuber & Manstead, 2009; Woodzicka, 2008). These findings suggested the likelihood of test-taker 6's assessment on the face-to-face test to be more positive; contrariwise, the scores were lower. One possible argument is that the Duchenne smiles generated by that test-taker did not represent feelings of confidence but were rather the result of negative feelings such as uncertainty. This argument is supported by the literature. Previous research has shown that Duchenne smiles can be produced when individuals are hiding negative emotions (Gunnery, Hall, & Ruben, 2013). Based on these results, I postulate that the high frequency of production of both non-Duchenne and Duchenne smiles might have influenced the perception and, possibly, the assessment of that test-taker during the face-to-face speaking test. This argument is corroborated by previous research which has indicated that different smiles are perceived differently by interlocutors and may affect the decoding of a message (Krys et al., 2014; Lau, 1982; Matsumoto & Kudoh, 1993).

Although the Duchenne smile has been shown to positively influence individuals' perceptions of speakers in North America, differences between Duchenne and non-Duchenne smiles may be tenuous in some cultures, and discriminating between both types may require examiners to be aware of such differences. For instance, research has shown that individuals from Western countries and those from China evaluate Duchenne smiles differently (Thibault et al., 2012). Moreover, previous studies have shown that although Duchenne smiles have been shown to positively influence how individuals are perceived in North America, intelligence, honesty, and sincerity are associated with non-smiling faces in Japan (Matsumoto & Kudoh, 1993). Based on the findings from this research, it is not possible to suggest that examiners were aware of possible cultural differences in the production of smiles; in actuality, comments about "natural" smiles suggest that examiners perceived smiles according to their personal, culturally-influenced

perception and judgment.

5.1.2 Scores per criterion

An examination of the intra-rater agreement rate showed that despite the high agreement rate, differences were found in the scores assigned to the each test-taker's performance in the face-to-face and the audio-recorded tests. The overall assessment of test-takers' performances in the face-to-face and audio-recorded speaking tests revealed that two test-takers received higher scores in the face-to-face tests and two in the audio-recorded tests. These findings contrast with previous research that has suggested that examinees consistently receive higher scores during face-to-face assessments when scores are compared to those assigned to the audio-recordings of the same tests (Nambiar & Goon, 1993), which prompted the investigation of how each criterion was assessed. Results revealed that some criteria were more positively assessed during the face-to-face speaking tests.

Findings show that the assessment of test-takers' *fluency and coherence* consistently resulted in higher scores being assigned to the performances in the face-to-face speaking tests. Similar findings were observed during the analysis of the scores assigned to the criterion *pronunciation*, in which three out of four test-takers (1, 4, and 5) received higher scores (between 0.5 and 1.0) on their performances in the face-to-face speaking tests. These findings are supported by the literature. Nonverbal cues have also been shown to enhance content understanding (Beattie & Shovelton, 2005; Berger & Popelka, 1971; Graham & Argyle, 1975; Riseborough, 1981). They may also facilitate negotiation of meaning, help speakers to recall lexical items faster, and are especially effective when the intelligibility of the speech is reduced (Hadar, Wenkert-Olenik, Krauss, & Soroket, 1998). Moreover, the ability to use nonverbal cues appropriately in a second language has been shown to be an intrinsic part of language fluency

(Saville-Troike, 2003).

Although one might suggest that visual input could have positively influenced the examiners' assessment of the test-takers' performances in these two criteria, questions remained as to why test-taker 6 received a lower score for her pronunciation skills in the face-to-face speaking test. In an attempt to find an answer to that question, I referred to the qualitative data; review of the data revealed, as previously mentioned, that examiner F reported not only having difficulty understanding the accent of CL1 speakers, but also stated a preference for no visual input during the assessment of test-takers' speaking skills. Such preference, in addition to the examiner's account of being distracted by the test-taker's hand gesture, might explain the lower score assigned to the test-taker's performance in the face-to-face speaking test. It is possible to argue, therefore, that the lower score test-taker 6 received in the criterion *pronunciation* could have been influenced by the examiner's difficulty understanding CL1 speakers' accent. Based on these findings, I postulate that visual input, the only element that differed in the two sets of assessed performances, may have been a factor in the assessment of the two criteria during the face-to-face speaking tests.

In regards to *grammatical range and accuracy*, findings show no differences in the assessment of the performances of test-takers 1 and 5 in the face-to-face and audio-recorded tests; however, results from the evaluations of the performances of test-takers 4 and 6 showed that they were assigned lower scores during the face-to-face speaking tests. To investigate if other factors could have influenced the assessment of these test-takers' performances, I referred to the qualitative data. The results from the data analysis did not indicate if the scores assigned could have been influenced by the nonverbal cues. These findings were therefore deemed inconclusive and require further investigation.

Results from the analysis of the scores assigned to *lexical resource* showed that two test-

takers received higher scores during the face-to-face speaking tests (test-takers 1 and 5) and two received lower scores (test-takers 4 and 6). The analysis of the lexical range of test-takers' performances showed that test-taker 5 produced the most K1 words of all test-takers (89.41% of all the words she used in the test). Findings from the qualitative analysis, showed that examiner E, who conducted the face-to-face test, did not refer to the test-taker's lexical range during the recall session. These findings differ from those from the analysis of the assessment of test-taker 6's performance. This test-taker produced the highest number of AWL words and, as described by the examiner, seemed to be someone who possessed the vocabulary of an individual with knowledge about the topic being discussed. This initial observation was further supported by the examiner, who mentioned that the test-taker knew technical words and that, based on the vocabulary used, the test-taker might be familiar with a subject related to the topic. The examiner also mentioned that the test-taker's gestures seemed to indicate her considerable knowledge of the topic being discussed. Despite these results, the score assigned to her face-to-face performance was lower than the mean score assigned to her performance in the audio-recorded test. Review of the findings from the qualitative analysis and from the analysis of the scores assigned and the generated nonverbal cues did not provide an answer or explanation for these results. The overall lack of identifiable patterns in the data regarding this criterion deemed these findings inconclusive.

5.1.3 Examiners' perceptions of nonverbal cues vis-à-vis test-takers' performances.

Findings from an analysis of the examiners' assessments of the test-takers' performances in the face-to-face speaking tests in relation to the examiners' perceptions of test-takers' nonverbal cues indicate that examiners were mostly aware of the nonverbal cues being produced during the tests. Individually, three out of four examiners shared thoughts about how they perceived and

decoded test-takers' cues.

To investigate if the examiners' perceptions of the nonverbal cues could be related to how each dyad interacted during the speaking test, I examined their accounts of nonverbal encouragement and mimicked cues that were identified in the data. Findings from the analysis of the role of mimicry of nonverbal cues as a form of encouragement showed that the examiners who mimicked test-takers' nonverbal cues more frequently (examiners D and E) also assigned the highest scores to their test-takers during the face-to-face speaking tests (test-takers 4 and 5). In regards to the dyad examiner F/test-taker 6, the examiner mimicked the least number of cues generated (1.36%), and the test-taker was also the one who received the lowest overall scores in the face-to-face speaking test. These findings are supported by previous studies. Being mimicked has been shown to lead individuals to have a more favourable attitude towards individuals who mimic them (Fischer-Lokou et al., 2011), which might have been reflected on what examiners D and E perceived as the 'natural,' 'relaxed attitudes of test-takers 4 and 5, respectively. Research on rapport and mimicry has also shown that the two are positively correlated. For instance, counselors who mimicked the nonverbal cues of clients perceived the clients to be warmer, more genuine, and to possess greater expertise (Chartrand & Dalton, 2007). Although it is possible that examiners D and E mimicked the nonverbal cues generated by test-takers 4 and 5 because they perceived the test-takers to be strong students, findings from previous studies have suggested that mimicry may lead to greater rapport and to individuals being assessed more positively (Chartrand & Dalton, 2007). One possible argument, therefore, would be that the examiners mimicked test-takers' nonverbal cues in an attempt to show engagement and provide encouragement; research has also indicated that mimicry (and consequent affiliation) may be a strategy of social coordination (Chartrand & Bargh, 1999; Stel et al., 2005).

In summary, findings from a comparative analysis of the scores assigned to each criterion during the face-to-face and audio-recorded speaking tests suggest that nonverbal cues may have played a role in the assessment of test-takers' speaking skills. More specifically, nonverbal cues may have influenced the assessment of test-takers' *fluency and coherence* and *pronunciation*. Additionally, findings indicate that the lowest scoring test-takers used the most nonverbal cues (non-Duchenne smiles, adaptor gestures, high frequency of eye contact) that have been shown to have a possible negative influence on how individuals are perceived, while a confluence of nonverbal cues, which have been shown to positively influence the perception of speakers (e.g., as eye contact, Duchenne smiles, and illustrators), might have compensated for the lower lexical range of test-taker 5 and positively influenced how she was perceived by examiner E.

5.2 Implications

5.2.1. Methodological Implications.

The combination of the well-established research methods used in this study to investigate the role of nonverbal cues in second language assessment can be considered new.

First, this study, to my knowledge, is the first to investigate the role of nonverbal cues produced by EAL test-takers on scores assigned by EL1 examiners during simulated IELTS Speaking Tests. Tests were video-recorded to collect visual data for the analysis of the nonverbal cues produced during the test and audio-recorded to examine how test-takers performances were evaluated without the influence of visual input. To better understand how scores were assigned, examiners were invited to take part in recall sessions, in which they had the opportunity to share thoughts they recalled from the face-to-face speaking tests.

Second, the integration of quantitative and qualitative research methods provided an opportunity for an in-depth analysis of the data. Results from the qualitative analysis provided an

insight into how examiners perceived test-takers' nonverbal behaviour and also into their decision-making processes during the tests. These results were combined with the findings of the quantitative analysis, which provided information on scores assigned to test-takers both in the face-to-face and audio recorded performances as well as on the frequency of generated nonverbal cues. Unlike previous studies on nonverbal cues (DeGroot & Gooty, 2009), nonverbal and verbal data were integrated for the analysis of the nonverbal cues, reflecting the true nature of communicative events. Finally, the integration of the data collected from the different sources (video recording, questionnaires, rating sheets, and recall sessions) provided a more detailed picture of how examiners evaluated test-takers' performances in the speaking tests.

5.2.2 Pedagogical and Practical Implications

The findings from this study suggest that nonverbal cues may play a role in the assessment of test-takers' performances during face-to-face speaking tests. These findings are partially supported by previous studies on the role of nonverbal behaviour on the perception and evaluation of speakers (Briegel-Jones, 2014).

Instructors might want to consider including information on culture-specific nonverbal communication in their practice. Fluency in a second language is made of more than just linguistic fluency or knowledge of cultural elements such as holidays or history (Saville-Troike, 2003). Research has also shown that individuals that generate nonverbal cues similar to those produced by other individuals are perceived more positively (Jenkins & Parra, 2003). Learning about the use of non-linguistic elements in intercultural exchanges may increase EAL speakers' confidence level as such knowledge may provide information that could facilitate exchanges with EL1 speakers. EAL test-takers may also benefit from these findings; results may help raise test-takers' awareness that inaccurate perception of nonverbal cues produced by them may lead to

scores that do not reflect their competence.

Examiners should also be made aware that nonverbal cues might play a role during the assessment of test-takers' performances. Although the decoding process is mostly unconscious, results from this study showed that examiners were aware of some of the nonverbal cues that were being generated by the test-takers. Examiners may benefit from training sessions that focus on raising awareness of differences in the production of nonverbal cues and how those differences may influence the overall assessment of test-takers' performances.

Nonverbal cues are an intrinsic part of communication, and test-designers might want to consider their possible role on the assessment of test-takers' language proficiency. One suggestion would be incorporating information regarding nonverbal communication to examiners' training material. This information may help raise examiners' awareness of the possible influence of nonverbal cues on scores assigned. A previous study by Briegel-Jones (2014) has indicated that nonverbal cues generated by examiners during standardized language proficiency speaking tests might affect the performance of test-takers; due to the interactional nature of communication, it is most likely that this influence goes both ways. While it is not being suggested that examiners should learn all the culturally-influenced nonverbal cues, it might be valuable to offer examiners the opportunity to learn more about how nonverbal communication might play a role on the assessment of speakers' communicative competence.

5.3 Limitations

5.3.1. Study Design.

The small (non-probabilistic) sample of IELTS test-takers of this exploratory study is a limitation to this study. Findings, therefore, are not conclusive but rather suggestive of the

possible role of the three nonverbal cues investigated on the assessment of test-takers' speaking skills. To validate these findings, studies with larger sample pools should be conducted as they might provide additional evidence that could be generalizable to other contexts (other cultural groups, a larger number of individuals, and so on).

5.3.2. Setting.

The study's setting and the presence of the researcher in the room should be seen as possible limitations. The lab in which tests took place was not designed specifically for testing conditions, which might have affected participants' performances. Moreover, the presence of the researcher in the room could have influenced the performance of both test-takers and examiners despite careful consideration being taken in regards to the placement of the data collection equipment and to the seating arrangements during data collection sessions.

5.3.3 IELTS Speaking Test and Examiners.

The use of a simulated IELTS Speaking Test, which may be considered a limitation, resulted from the impossibility of an official test being used in the study. While the original design of the study intended to include certified IELTS examiners, changes to their contract with IELTS prevented their participation without prior approval from IELTS. Although three out of four examiners did not have prior IELTS training or experience as a certified IELTS examiner, the use of experienced examiners who were familiar with the test was intended to minimize this limitation.

5.3.4. Video-stimulated Verbal Recall Sessions.

Accuracy of reports gathered from video-stimulated verbal recall sessions has been questioned (Gass & McKay, 2000). To minimize issues regarding the ability of examiners to

recall their thoughts during the tests, the recall sessions were conducted immediately after the conclusion of each speaking test. Furthermore, examiners had unlimited time to speak, and the questions asked during the sessions were designed to help examiners verbalize their recalled thoughts. While efforts were made to minimize issues, it is important to consider that recall is never truly complete.

5.3.5 Data Analysis.

First, it is important to acknowledge that subjectivity cannot be completely avoided in the analysis of qualitative data. Conversely, the nature of qualitative analyses makes it possible for an in-depth examination of the data to be conducted. The method and procedures of this study were designed with the intention to minimize possible analytical biases. Although this study intended to include member check to strengthen its findings, that was not feasible as two examiners were not reachable during the last stages of the study. Having come from a cultural background that differs from the test-takers', it is also possible that my cultural biases may have influenced the analysis of the nonverbal cues that were produced during the face-to-face speaking tests.

Second, the exclusion of three out of the four criteria that are assessed in the IELTS Speaking Test from the data analysis may be considered a limitation. Having examined every criterion vis-à-vis the scores assigned and the nonverbal cues generated could have resulted in richer findings. However, the lack of access to precise information regarding the assessment of those criteria led to the decision to have them excluded from this study. Moreover, although the investigation of several variables was conducted, this study did not examine the strength of the relationship among the variables.

5.3.6 Nonverbal Cues

Inclusion of all the nonverbal cues generated by the test-takers could have resulted in much richer findings. Although the decision to focus on three cues out of the vast array that are commonly produced during face-to-face exchanges is a limitation, an analysis of all the test-takers' nonverbal behaviour was beyond the scope of this study.

It is important to also highlight that idiosyncratic differences could account for some of the variations identified in the test-takers' frequency of production of nonverbal cues. Although it is clear that visual input might have played a role on how test-takers' performances were assessed during the face-to-face speaking tests, this study makes no claim that nonverbal cues are solely responsible for that difference.

Finally, the coding of nonverbal cues by a single coder might be considered a limitation. Although it would have been valuable to have the three nonverbal cues (and their subcategories) coded by a second coder, time constraints made that measure invariable.

5.4 Future Research Directions

This study intended to develop an initial understanding of the role of smiles, eye contact, and gestures on the assessment of test-takers' speaking skills through a detailed exploratory analysis. Despite its limitations, knowledge acquired from the in-depth analysis of the data has provided valuable information about how these cues might have played a role in the assessment of the test-takers' performances in this still rather unexplored context. If cultural differences in how nonverbal cues are encoded and decoded play a role on how individuals are evaluated, results from the assessment of test-takers' performances may not reflect their communicative competence.

The findings from this study have provided an initial insight into the interplay between nonverbal cues and the assessment of test-takers' performances in face-to-face speaking tests. Further investigation with a larger pool of participants should be considered as that might result in findings that are generalizable. One possible direction would be to investigate the effectiveness of a training session focusing on raising examiners' sensitivity to cultural differences in the production of nonverbal cues and their possible role in relation to the assessment of test-takers' performances. Researchers should also consider designing a mixed methods study to investigate the strength of the relationship of the variables examined in this study.

Future studies should also consider investigating mimicry of examiners' nonverbal cues as well as test-takers' as results may provide a better understanding of the role nonverbal behaviour on rapport building between the interlocutors and on the assessment of test-takers' performances in face-to-face speaking tests.

Finally, the investigation of other nonverbal cues generated by test-takers should be considered in future studies. Inclusion of cues such as regulator gestures and head nods, which have been shown to play a role in turn-taking and help regulate the flow of exchanges, may help paint a clearer picture of how the nonverbal interaction between EL1 examiners and EAL test-takers develops throughout the speaking test and how such interaction may affect the assessment of test-takers' performances.

5.5. Conclusion

This study, to my knowledge, is the first one to examine the role of smiles, gestures, and eye contact in the assessment of EAL test-takers' performance during simulated IELTS Speaking Tests. Although further investigation is necessary to examine the possible connection between nonverbal cues and scores assigned, the findings presented here provide a valuable contribution

to the fields of second language assessment and intercultural communication. Nonverbal behaviour, an intrinsic part of communication, is often produced and decoded unconsciously. The number of studies that have investigated the role of nonverbal communication in second language assessments does not reflect its vital role in how messages are decoded and how individuals are perceived. Knowledge acquired from this study may help equip EL2 speakers with information that could help increase their confidence level not only within the speaking test context but also during interactions with EL1 speakers.

On a personal note, I believe that this study has provided me with an invaluable opportunity to learn more about myself as a second language speaker. The knowledge that I have acquired from designing and implementing the study and from the data analysis has increased my overall awareness of the nonverbal cues I generate and how others may perceive them. Observing the interplay between verbal and nonverbal communication in intercultural exchanges has led me to reflect on the many elements that individuals may unconsciously assess during encounters with individuals from different cultural backgrounds. I believe that sharing this knowledge may help other second language speakers to become more aware of the fact that elements other than verbal exchanges may affect how information is perceived.

References

- Adams Jr., R. B., Nelson, A. J., & Purring, K. (2013). Eye behaviour. In J. A. Hall & M. K. Knapp (Eds.), *Nonverbal communication* (pp. 229-261). Berlin, Germany: De Gryter.
- Albright, L., Kennedy, D. A., & Malloy, T. E. (1988). Consensus in personality judgments at zero acquaintance. *Journal of Personality and Social Psychology*, 55(3), 387-395.
- Alderson, J. C., & Banerjee, J. (2002). Language testing and assessment (part 2). *Language Teaching*, 35(2), 70-113.
- Ambady, N., Hallahan, M., & Rosenthal, R. (1995). On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and Social Psychology*, 69(3), 518-529.
- Argyle, M. (1970). Eye-contact and distance: A reply to Stephenson and Rutter. *British Journal of Psychology*, 61(3), 395-396.
- AUCC (2011). *Trends in higher education, vol. 1 – Enrolment* [Data file]. Retrieved from <http://www.cais.ca/uploaded/trends-2011-vol1-enrolment-e.pdf>.
- Bachman, L. F. (1988). Problems in examining the validity of the ACTFL oral proficiency interview. *Studies in Second Language Acquisition*, 10(02), 149-164.
- Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford, UK: OUP.
- Bagarić, V. (2007). Defining communicative competence. *Metodika*, 8(1), 94-103.
- Barrick, M. R., Swider, B. W., & Stewart, G. L. (2010). Initial evaluations in the interview: Relationships with subsequent interviewer evaluations and employment offers. *Journal of Applied Psychology*, 95(6), 1163-1172.
- Bavelas, J., Kenwood, C., Johnson, T., & Phillips, B. (2002). An experimental study of when and how speakers use gestures to communicate. *Gesture*, 2(1), 1-17.

- Beattie, G., & Shovelton, H. (2005). Why the spontaneous images created by the hands during talk can help make TV advertisements more effective. *British Journal of Psychology*, 96(1), 21-37.
- Beattie, G., Webster, K., & Ross, J. (2010). The fixation and processing of the iconic gestures that accompany talk. *Journal of Language and Social Psychology*, 29(2), 194-213.
- Berger, K. W., & Popelka, G. R. (1971). Extra-facial gestures in relation to speechreading. *Journal of Communication Disorders*, 3(4), 302-308.
- Boyatzis, R. E. (1998). *Transforming qualitative information: Thematic analysis and code development*. Sage.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- Briegel-Jones, L. (2014). An investigation into nonverbal behaviour in the oral proficiency interview: Perceptions of interviewer variability and the impact on candidates (unpublished Master's thesis). Newcastle University, UK.
- Brown, A. (2000). An investigation of the rating process in the IELTS oral interview. *IELTS Research Reports*, 3, 48-84.
- Brown, A. (2006a). An examination of the rating process in the revised IELTS Speaking Test. *IELTS Research Reports*, 6, 1-30.
- Brown, A. (2006b). Candidate discourse in the revised IELTS Speaking Test. *IELTS Research Reports*, 6, 71-90.
- Buck, R. (1988). Nonverbal communication: Spontaneous and symbolic aspects. *American Behavioral Scientist*, 31(3), 341-354.
- Buck, R. W., & Powers, S. R. (2013). Encoding and display. In J. A. Hall & M. K. Knapp (Eds.), *Nonverbal Communication* (pp. 403-440). Berlin, Germany: De Gruyter Mouton.

- Bull, P. & Doody, J. P. (2013). Gesture and body movement. In J. A. Hall & M. K. Knapp (Eds.), *Nonverbal communication* (pp. 205-227). Berlin, Germany: De Gruyter Mouton.
- Burgoon, J. K., Buller, D. B., & Woodall, W. G. (1989). *Nonverbal communication: The unspoken dialogue*. New York, NY: Harper & Row.
- Burgoon, J. K., & Poire, B. A. L. (1999). Nonverbal cues and interpersonal judgments: Participant and observer perceptions of intimacy, dominance, composure, and formality. *Communication Monographs*, 66(2), 105-124.
- Campbell, D. T., & LeVine, R. A. (1961). A proposal for cooperative cross-cultural research on ethnocentrism. *Journal of Conflict Resolution*, 5(1), 82-108.
- Canadian Bureau for International Education (CBIE), (2015). Canada's performance and potential in international education 2016. Retrieved from <http://cbie.ca>.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
- Cassell, J. (1998). A framework for gesture generation and interpretation. *Computer vision in human-machine interaction*, 191-215.
- Chalhoub-Deville, M., & Turner, C. E. (2000). What to look for in ESL admission tests: Cambridge certificate exams, IELTS, and TOEFL. *System*, 28, 523-539.
- Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception-behavior link and social interaction. *Journal of Personality and Social Psychology*, 76(6), 893-910.
- Chartrand, T. L., & Dalton, A. N. (2007). Mimicry. In R. F. Baumeister & K. D. Vohs, *Encyclopedia of Social Psychology* (pp. 572). Thousand Oaks, CA: Sage.
- Chen, C., Chen, H., & Lin, Y. (2013). The boundaries of effects on the relationship between interviewer moods and hiring recommendation. *Applied Psychology*, 62(4), 678-700.
- Csibra, G., & Gergely, G. (2006). Social learning and social cognition: The case for pedagogy. In

- Y. Munakata & M.H. Johnson (Eds.), *Processes of Change in Brain and Cognitive Development. Attention and Performance, Vol. 21.* (pp.249-274). Oxford, UK: Oxford University Press.
- Cobb, T., & Horst, M. (2011). Does word coach coach words. *Calico Journal*, 28(3), 639-661.
- Darwin, C., 1809-1882, & Barker, D. W. (1872). *The expression of the emotions in man and animals.* London, UK: J. Murray.
- DeGroot, T., & Gooty, J. (2009). Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Journal of Business and Psychology*, 24(2), 179-192.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111(2), 203-243.
- Duchenne, C.-B. (1862). The mechanism of human facial expression. *Jules Renard.*
- Duchenne, G.-B., & Cuthbertson, R. A. (1990). *The mechanism of human facial expression: Mécanisme de la physionomie humaine.* Cambridge, UK: Cambridge University Press.
- Dunbar, N. E., & Burgoon, J. K. (2005). Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships*, 22(2), 207-233.
- Duncan, S., & Fiske, D. W. (1977). *Face-to-face interaction: Research, methods, and theory.* Hillsdale, NY: L. Erlbaum Associates.
- Durant, A., & Shepherd, I. (2009) 'Culture' and 'communication' in intercultural communication. *European Journal of English Studies*, 13(2), 147-162.
- Efran, J. S. (1968). Looking for approval: Effects on visual behavior of approbation from persons differing in importance. *Journal of Personality and Social Psychology*, 10(1), 21-25.
- Efron, D. (1941). *Gesture and environment.* Oxford, UK: King Crown's Press.
- Ekman, P. (1976). Movements with precise meanings. *Journal of Communication*, 26(3), 14-26.

- Ekman, P. (2004). Emotional and conversational nonverbal signals. In J. M. Larrazabal & L.A. Pérez (Eds.), *Language, knowledge, and representation* (pp. 39-50). Netherlands: Kluwer Academic Publishers.
- Ekman, P., & Friesen, W. V. (1969). Nonverbal leakage and clues to deception. *Psychiatry*, 32(1), 88-106.
- Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior - categories, origins, usage and coding. *Semiotica*, 1(1), 49-98.
- Ekman, P., & Friesen, W. V. (1974). Detecting deception from body or face. *Journal of Personality and Social Psychology*, 29(3), 288-298.
- Ekman, P., & Friesen, W. V. (1977). Nonverbal behavior. In Ostwald, P. F. (Ed.), *Communication and social interaction* (pp. 37-46). New York, NY: Grune & Stratton.
- Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. New York, NY: Oxford University Press.
- Emerson, R. W. (1883). *Letters and social aims*. Boston, MA: Houghton, Mifflin and Company.
- Etcoff, N. L., & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition*, 44(3), 227-240.
- Exline, R. V., & Fehr, B. J. (1982). The assessment of gaze. In K. R. Scherer & P. Ekman (Eds.), *Handbook of methods in nonverbal behavior research* (91-135). Cambridge, UK: Cambridge University Press.
- Exline, R. V., Thibaut, J., Hickey, C. B., & Gumpert, P. (1970). Visual interaction in relation to machiavellianism and an unethical act. In R. Christie & F. L. Geis (Eds.). *Studies in machiavellianism* (pp. 53-75). New York, NY: Academic Press.
- Fischer-Lokou, J., Martin, A., Guéguen, N., & Lamy, L. (2011) Mimicry and propagation of

- prosocial behaviour in a natural setting. *Psychological Reports*, 108(2), 599-605.
- Fiksdal, S. (1988). Verbal and nonverbal strategies of rapport in cross-cultural interviews. *Linguistics and Education*, 1(1), 3-17.
- Friesen, W. V. (1973). *Cultural differences in facial expressions in a social situation: An experimental test on the concept of display rules* (Doctoral dissertation, ProQuest Information & Learning).
- Friesen, W. V., Ekman, P., & Wallbott, H. (1979). Measuring hand movements. *Journal of Nonverbal Behavior*, 4(2), 97-112.
- Fujiwara, K., & Daibo, I. (2014). The extraction of nonverbal behaviors: Using video images and speech-signal analysis in dyadic conversation. *Journal of Nonverbal Behavior*, 38(3), 377-388.
- Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors. *System*, 32(3), 301-319.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: L. Erlbaum Associates.
- Gifford, R. (1994). A lens-mapping framework for understanding the encoding and decoding of interpersonal dispositions in nonverbal behavior. *Journal of Personality and Social Psychology*, 66(2), 398-412.
- Gifford, R. (2006). Personality and nonverbal behavior. In V. Manusov & M. L. Patterson (Eds.), *The Sage Handbook of Nonverbal Behavior*, (pp. 201-218). Thousand Oaks, CA: Sage.
- Gifford, R., Ng, C. F., & Wilkinson, M. (1985). Nonverbal cues in the employment interview: Links between applicant qualities and interviewer judgments. *Journal of Applied Psychology*, 70(4), 729-736.
- Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal

- meaning by gestures (1). *International Journal of Psychology*, 10(1), 57-67.
- Gregersen, T. S. (2005). Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals*, 38(3), 388-400.
- Gudykunst, W. B., & Ting-Toomey, S. (1988). Culture and affective communication. *American Behavioral Scientist*, 31(3), 384-400.
- Gullberg, M. (2010). Methodological reflections on gesture analysis in second language acquisition and bilingualism research. *Second Language Research*, 26(1), 75-102.
- Gullberg, M. (2006). Some reasons for studying gesture and second language acquisition (hommage à Adam Kendon). *IRAL - International Review of Applied Linguistics in Language Teaching*, 44(2), 103-124.
- Gunnery, S. D., & Hall, J. A. (2014). The Duchenne smile and persuasion. *Journal of Nonverbal Behavior*, 38(2), 181-194.
- Gunnery, S. R., Hall, J. A., & Ruben, M. A. (2013). The deliberate Duchenne smile: Individual differences in expressive control. *Journal of Nonverbal Behavior*, 37(1), 29-41.
- Gunnery, S. D., & Ruben, M. A. (2016). Perceptions of Duchenne and non-Duchenne smiles: A meta-analysis. *Cognition and Emotion*, 30(3), 501-515.
- Hadar, U., Wenkert-Olenik, D., Krauss, R., & Soroker, N. (1998). Gesture and the processing of speech: Neuropsychological evidence. *Brain and language*, 62(1), 107-126.
- Hager, J. C., Ekman, P., & Friesen, W. V. (2002). *Facial action coding system*. Salt Lake City, UT: A Human Face.
- Hall, J. A. (1978). Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 85(4), 845-857.
- Hall, J. A. (1984). *Nonverbal sex differences: Communication accuracy and expressive style*. Baltimore, MD: The Johns Hopkins University Press.

- Hall, J. (2004). Cicero and Quintilian on the oratorical use of hand gestures. *Classical Quarterly*, 54(1), 143-160.
- Hall, N. R., Millings, A., & Bouças, S. B. (2012). Adult attachment orientation and implicit behavioral mimicry. *Journal of Nonverbal Behavior*, 36(4), 235-247.
- Harrigan, J. A. (2013). Methodology: Coding and studying nonverbal behavior. In J. A. Hall & M. K. Knapp (Eds.), *Nonverbal Communication* (pp. 35-68). Berlin, Germany: De Gruyter Mouton.
- He, L. (2012). Topical knowledge and ESL writing. *Language Testing*, 29(3), 443-464.
- Henningsen, D. D., Valde, K. S., & Davies, E. (2005). Exploring the effect of verbal and nonverbal cues on perceptions of deception. *Communication Quarterly*, 53(3), 359-375.
- Hess, U., Kappas, A., McHugo, G. J., Kleck, R. E., & Lanzetta, J. T. (1989). An analysis of the encoding and decoding of spontaneous and posed smiles: The use of facial electromyography. *Journal of Nonverbal Behavior*, 13(2), 121-137.
- Hess, U., & Kleck, R. E. (1994). The cues decoders use to in attempting to differentiate emotion-elicited and posed facial expressions. *European Journal of Social Psychology*, 24(3), 367-381.
- Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, 36(2), 153-166.
- Hirsbrunner, H. P., Frey, S., & Crawford, R. (1987). Movement in human interaction: Parameter formation and analysis. In A. W. Siegmann & S. Feldstein (Eds.), *Nonverbal behavior and communication* (pp. 99-140). Hillsdale, NJ: Erlbaum.
- Hirvela, A., Nussbaum, A., & Pierson, H. (2012). ESL students' attitudes toward punctuation. *System*, 40(1), 11-23.
- Holler, J., & Wilkin, K. (2011). Co-speech gesture mimicry in the process of collaborative

- referring during face-to-face dialogue. *Journal of Nonverbal Behavior*, 35(2), 133-153.
- Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior*, 33(2), 73–88.
- House, T. (2001). Equal hiring, a must. *Computerworld*, 35, 32-34.
- Huang, L.-S. (2011). Are we having the effect we want? Implementing outcomes assessment in an academic English language-support unit using a multi-component approach. *WPA Journal*, 35(1), 11-45.
- Huang, L.-S. (2014). *Cognitive processes involved in performing the IELTS Speaking Test: Respondents' strategic behaviours in simulated testing and non-testing contexts (IELTS Research Report Series No. 1)*. Adelaide, South Australia, Australia: IDP-IELTS, 1-51.
- Huang, L.-S. (2013). Video-stimulated verbal recall. *Sage Cases in Methodology*.
- Hudelson, S. (1989). *Write on: Children writing in ESL*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, 81(5), 459-473.
- Hymes, D. (1972). Competence and performance in linguistic theory. In R. Huxley and E. Ingram (Eds.), *Language acquisition: Models and methods* (pp. 3-24). London: Academic Press.
- IELTS. (2016). About IELTS. Retrieved from http://www.ielts.org/about_us.aspx
- IELTS. (2016). IELTS – Test-takers – Speaking sample. Retrieved from http://www.ielts.org/test_takers_information/test_sample/speaking_sample.aspx
- Imada, A. S., & Hakel, M. D. (1977). Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Journal of Applied*

Psychology, 62(3), 295-300..

- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29(1), 24-4.
- Jenkins, S., & Parra, I. (2003). Multiple layers of meaning in an oral proficiency test: The complementary roles of nonverbal, paralinguistic, and verbal behaviors in assessment decisions. *The Modern Language Journal*, 87(1), 90-107.
- Jokinen, K., & Allwood, J. (2010). Hesitation in Intercultural Communication: Some observations on Interpreting Shoulder Shrugging. In *Proceedings of the international workshop on agents in cultural context, The first international conference on culture and computing*, 25-37.
- Johnston, L., Miles, L. K., & Macrae, C. N. (2010). Why are you smiling at me? Social functions of enjoyment and nonenjoyment smiles. *British Journal of Social Psychology*, 49(1), 107-127.
- Kang, O. (2012). Impact of rater characteristics and prosodic features of speaker accentedness on ratings of international teaching assistants' oral performance. *Language Assessment Quarterly*, 9, 249-269.
- Kappas, A., Krumhuber, E., & Küster, D. (2013). Facial behavior. In J. A. Hall & M. K. Knapp (Eds.), *Nonverbal Communication* (pp. 129-165). Berlin, Germany: De Gruyter Mouton.
- Kelly, S. D., Barr, D. J., Church, R. B., & Lynch, K. (1999). Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language*, 40(4), 577-592.
- Kelly, S. D., & Goldsmith, L. H. (2004). Gesture and right hemisphere involvement in evaluating lecture material. *Gesture*, 4(1), 25-42.
- Kendon, A. (1970). Movement coordination in social interaction: Some examples described. *Acta*

Psychologica, 32(1), 101-125.

Kendon, A. (1997). Gesture. *Annual Review of Anthropology*, 26(1), 109-128.

Kendon, A. (2000). Language and gesture: Unity or duality. In D. McNeill (Ed.), *Language and gesture* (pp. 47–63). New York, NY: Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Kenny, D. A., Horner, C., Kashy, D. A., & Chu, L. C. (1992). Consensus at zero-acquaintance: Replication behavioral cues, and stability. *Journal of Personality and Social Psychology*, 62(1), 88-97.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, 4(5), 900-913.

Kleinke, C. L. (1986). Gaze and eye contact: A research review. *Psychological Bulletin*, 100(1), 78-100.

Knapp, M. L., & Daly, J. A. (2002). *Handbook of interpersonal communication*. Thousand Oaks, CA: Sage.

Knapp, M. L., & Hall, J. A. (2006). *Nonverbal communication in human interaction*, 6th ed. Belmont, CA: Thomson Wadsworth.

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in experimental social psychology*, 28, 389-450.

Kraut, R. E., & Johnston, R. E. (1979). Social and emotional messages of smiling: An ethological approach. *Journal of Personality and Social Psychology*, 37(9), 1539-1553.

Krumhuber, E. G., & Manstead, A. S. R. (2009). Can Duchenne smiles be feigned? New

- evidence on felt and false smiles. *Emotion*, 9(6), 807–820.
- Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., & Rosin, P. L. (2009). Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting. *Journal of Nonverbal Behavior*, 33(1), 1-15.
- Krys, K., Hansen, K., Xing, C., Szarota, P., & Yang, M. (2014). Do only fools smile at strangers? Cultural differences in social perception of intelligence of smiling individuals. *Journal of Cross-Cultural Psychology*, 45(2), 314-321.
- Lakin, J. L., & Chartrand, T. L. (2003). Using nonconscious behavioral mimicry to create affiliation and rapport. *Psychological science*, 14(4), 334-339.
- Lakin, J. L., Chartrand, T. L., & Arkin, R. M. (2008). I am too just like you: Nonconscious mimicry as an automatic behavioral response to social exclusions. *Psychological Science*, 19(8), 816–822.
- Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, 27(3), 145-162.
- Lau, S. (1982). The effect of smiling on person perception. *The Journal of Social Psychology*, 117(1), 63-67.
- Lauer, M. J. (2005). Interpreting nonverbal communication in interviews. *Performance Improvement*, 44(1), 39-40.
- LeVine, R. A., & Campbell, D. T. (1972). *Ethnocentrism: Theories of conflict, ethnic attitudes, and group behavior*. New York, NY: Wiley.
- Lewis, C., Lovatt, P., & Kirk, E. (2015). Many hands make light work: The facilitative role of gesture in verbal improvisation. *Thinking Skills and Creativity*, 17, 149-157.
- Liddicoat, A. J. (2009). Communication as culturally contexted practice: A view from

- intercultural communication. *Australian Journal of Linguistics*, 29(1), 115-133.
- Liden, R. C., Martin, C. L., & Parsons, C. K. (1993). Interviewer and applicant behaviors in employment interviews. *Academy of Management*, 36(2), 372-386.
- Luoma, S. (2004). *Assessing speaking*. Cambridge, UK: Cambridge University Press.
- Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review*, 19(3), 203–218.
- Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.
- Mahl, G. F. (1968). Gestures and body movement in interviews. In J. Schlien (Ed.), *Research in psychotherapy*. Washington, DC: American Psychological Association.
- Maricchiolo, F., Gnisci, A., Bonaiuto, M., & Ficca, G. (2009). Effects of different types of hand gestures in persuasive speech on receivers' evaluations. *Language and Cognitive Processes*, 24(2), 239-266.
- Matsumoto, D., Hwang, H. S. (2013). Body and gestures. In D. Matsumoto, M. G. Frank, & H. S. Hwang (Eds.), *Nonverbal communication: Science and applications* (pp. 75-96). Thousand Oaks, CA: Sage.
- Matsumoto, D., & Kudoh, T. (1993). American-Japanese cultural differences in attributions of personality based on smiles. *Journal of Nonverbal Behavior*, 17(4), 231-243.
- McGovern, T. V., & Tinsley, H. E. A. (1978). Interviewer evaluations of interviewee nonverbal behaviour. *Journal of Vocational Behavior*, 13(2), 162-171.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought* (Paperback ed.). Chicago, IL: University of Chicago Press.
- Mehu, M., Little, A. C., & Dunbar, R. I. (2007). Duchenne smiles and the perception of generosity and sociability in faces. *Journal of Evolutionary Psychology*, 5(1), 183-196.

- Merrylees, B. (1999). An investigation of speaking test reliability with particular reference to examiner attitude to the speaking test format and candidate/examiner discourse produced. *IELTS Research Reports*, 2, 1-35.
- Mewald, C. (2015). Lexical range and communicative competence of learners in bilingual schools in lower Austria. *Global Education Review*, 2(2), 98-113.
- Morthorst, B., Krogh, J., Erlangsen, A., Alberdi, F., & Nordentoft, M. (2012). Effect of assertive outreach after suicide attempt in the AID (assertive intervention for deliberate self-harm) trial: Randomised controlled trial. *BMJ (Clinical Research Ed.)*, 345, e4972.
- Murphy, N. A., Lehrfeld, J. M., & Isaacowitz, D. M. (2010). Recognition of posed and spontaneous dynamic smiles in young and older adults. *Psychology and Aging*, 25(4), 811-821.
- Nambiar, M. K., & Goon, C. (1993). Assessment of oral skills: A comparison of scores obtained through audio recordings to those obtained through face-to-face evaluation. *RELC Journal*, 24(1), 15-31.
- Naylor, R. W. (2007). Nonverbal cues-based first impressions: Impression formation through exposure to static images. *Marketing Letters*, 18(3), 165-179.
- Neuliep, J. W., Hintz, S. M., & McCroskey, J. C. (2005). The influence of ethnocentrism in organizational context: Perceptions of interviewee and managerial attractiveness, credibility, and effectiveness. *Communication Quarterly*, 53(1), 41-56.
- Nguyen, L. S., Frauendorfer, D., Mast, M. S., & Gatica-Perez, D. (2014). Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Transactions on Multimedia*, 16(4), 1018-1031.
- Nishiyama, K. (2000). *Doing business with Japan: Successful strategies for intercultural communication*. University of Hawaii Press.

- O'Loughlin, K. (2000). The impact of gender in the IELTS oral interview. *IELTS Research Reports, 3*, 1-28.
- O'Sullivan B., & Lu, Y. (2002). The impact on candidate language of examiner deviation from a set interlocutor frame in the IELTS Speaking Test. *IELTS Research Reports 6*, 1-25.
- Parton, S. R., Siltanen, S. A., Hosman, L. A., & Langenderfer, J. (2002). Employment interview outcomes and speech style effects. *Journal of Language and Social Psychology, 21*(2), 144-161.
- Pocovnicu, A., & Vasilache, S. (2012). An interview-based assessment of the influence of ethnocentrism in business. *Review of International Comparative Management, 13*(3), 478-492.
- Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology, 55*(1), 1-81.
- Read, J., & Nation, P. (2002). An investigation of the lexical dimension of the IELTS Speaking Test. *IELTS Research Reports, 6*, 1-25.
- Richmond, V. P., & McCroskey, J. C. (2000). The impact of supervisor and subordinate immediacy on relational and organizational outcomes. *Communications Monographs, 67*(1), 85-95.
- Riseborough, M. G. (1981). Physiographic gestures as decoding facilitators: Three experiments exploring a neglected facet of communication. *Journal of Nonverbal Behavior, 5*(3), 172-183.
- Rosenthal, R. (1978). *Sensitivity to nonverbal communication: The PONS test*. Baltimore, MD: Johns Hopkins University Press.
- Rosip, J. C., & Hall, J. A. (2004). Knowledge of nonverbal cues, gender, and nonverbal decoding

- accuracy. *Journal of Nonverbal Behavior*, 28(4), 267-286.
- Samovar, L. A., & Porter, R. E. (2003). Understanding intercultural communication: An introduction and overview. In Samovar, L. A., & Porter, R. E. (Eds.), *Intercultural communication: A Reader* (pp. 6-17). Belmont, CA: Wadsworth.
- Samovar, L. A., & Porter, R. E., McDaniel, E. R., & Roy, C. S. (2013). *Communication between cultures*. Boston, MA: Wadsworth.
- Sapir, E. (1935). Communication. In *Encyclopaedia of the Social Sciences*, vol. 4 (pp. 78-81). New York, NY.
- Sapir, E., 1884-1939, & Mandelbaum, D. G., 1911-1987. (1951). *Selected writings in language, culture and personality*. Berkeley, CA: University of California Press.
- Sapir, E. (1949). *Selected writings in language: culture and personality (Double)*. University of California Press.
- Savignon, S. J. (1972) *Communicative Competence: An Experiment in Foreign Language Teaching*. Philadelphia, PA: The Centre for Curriculum Development.
- Saville-Troike, M., (2003). *The ethnography of communication: An introduction* (3rd ed.). Malden, MA: Blackwell.
- Schilbach, L., Wohlschlaeger, A. M., Kraemer, N. C., Newen, A., Shah, N. J., Fink, G. R., & Vogeley, K. (2006). Being with virtual others: Neural correlates of social interaction. *Neuropsychologia*, 44(5), 718-730.
- Seedhouse, P., & Harris, A. (2014). Topic development in the IELTS Speaking Test. *Melts Research Report*, 12, 1-56.
- Senju, A., & Johnson, M. H. (2009). The eye-contact effect: Mechanisms and development. *Trends in Cognitive Sciences*, 13(3), 127-134.
- Singelis, T. (1994). Nonverbal communication in intercultural interactions. In R. Brislin & T.

- Yoshida (Eds.), *Improving intercultural interactions* (pp. 268-294). Thousand Oaks, CA: Sage.
- Sodikoff, C. L., Firestone, I. J. M., & Kaplan, K.J. (1974). Distance matching and distance equilibrium in the interview dyad. *Personality and Social Psychology Bulletin, 1*, 243-245.
- Sorell, J., & Shin, D. (2007). The next step in concordance-based language learning: Constructing an online language learning resource for high-frequency vocabulary and collocations. *International Journal of Learning, 13*(12), 217-221.
- Stel, M., Blascovich, J., McGall, C., & Vonk, R. (2005). When mimicry makes it worse. Paper presented at the European Association for Experimental Social Psychology 14th General Meeting, Würzburg.
- Swami, V., & Furnham, A. (2012). An investigation of self-rated cues believed to influence the judgment of intelligence in a zero-acquaintance context. *Journal of Applied Social Psychology, 42*(8), 2064-2076.
- Tatler, B. W., Kirtley, C., MacDonald, R. G., Mitchell, K. M. A., & Savage, S. W. (2014). The active eye: Perspectives on eye movement research. In M. Horsley, M. Eliot, B. A. Knight, & R. Reilly (Eds.), *Current trends in eye tracking research* (pp. 3-16). Switzerland: Springer International Publishing Switzerland.
- Tessler, R., & Sushelsky, L. (1978). Effects of eye contact and social status on the perception of a job applicant in an employment interviewing situation. *Journal of Vocational Behavior, 13*(3), 338-347.
- Thibault, P., Levesque, M., Gosselin, P., & Hess, U. (2012). The Duchenne marker is not a universal signal of smile authenticity – but it can be learned! *Social Psychology, 43*(4), 215-221.
- Vrij, A. (1993). Credibility judgments of detectives: The impact of nonverbal behavior, social

- skills, and physical characteristics on impression formation. *Journal of Social Psychology*, *133*(5), 601-610.
- Washburn, P. V., & Hakel, M. D. (1973). Visual cues and verbal content as influences on impressions formed after simulated employment interviews. *Journal of Applied Psychology*, *58*(1), 137-141.
- Washington, M. C., Okoro, E. A., & Thomas, O. (2012). Intercultural communication in global business: An analysis of benefits and challenges. *International Business & Economics Research Journal*, *11*(2), 217-222.
- Wheeler, R. W., Baron, J. C., Michell, S., & Ginsburg, H. J. (1979). Eye contact and the perception of intelligence. *The Bulletin of the Psychonomic Society*, *13*(2), 101-102.
- Wiesner, W. H., & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology*, *61*(4), 275-290.
- Wilhelm, O., Hildebrandt, A., Manske, K., Schacht, A., & Sommer, W. (2014). Test battery for measuring the perception and recognition of facial expressions of emotion. *Frontiers in Psychology*, *5*, 404.
- Winke, P., Gass, S., & Myford, C. (2013). Rater's L2 background as a potential source of bias in rating oral performance. *Language Testing*, *30*(2), 231-252.
- Woo, H. S., & Prud'homme, C. (1999). Cultural characteristics prevalent in the Chinese negotiation process. *European Business Review*, *99*(5), 313-322.
- Woodzicka, J. A. (2008). Sex differences in self-awareness of smiling during a mock job interview. *Journal of Nonverbal Behavior*, *32*(2), 109-121.
- Yamamoto, K., & Suzuki, N. (2006). The effects of social interaction and personal relationships on facial expressions. *Journal of Nonverbal Behavior*, *30*(4), 167-179.

Young, J. W., So, Y., & Ockey, G. J. (2013). Guidelines for best test development practices to ensure validity and fairness for international English language proficiency assessments.

Educational Testing Service.

Young, R. (2002). Discourse approaches to oral language assessment. *Annual Review of Applied Linguistics*, 22, 243-262.

Zhu, Y., Yang, Z., & Yuan, B. (2013). Vision based hand gesture recognition. *International Conference on Service Sciences (ICSS)*, 260-265.

Appendices

Appendix A

Speaking Test

Test #5

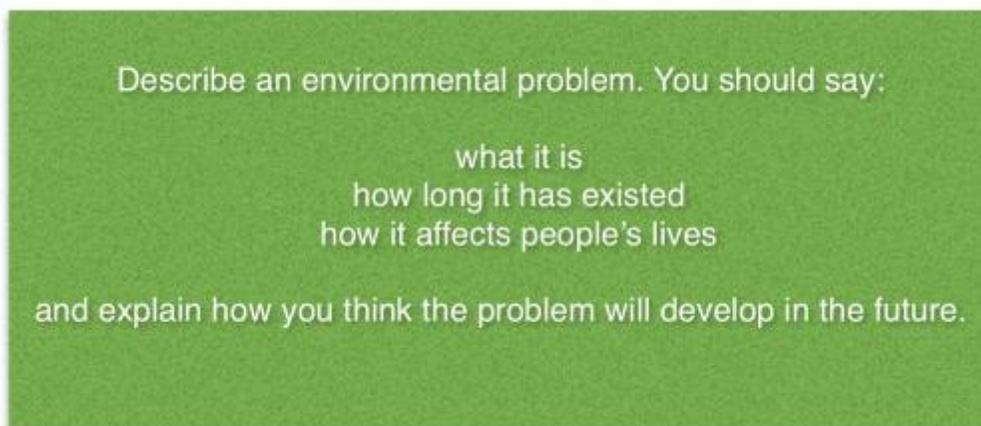
Part 1 - Introduction and interview (3 - 4 minutes)

[This part of the test begins with the examiner introducing himself or herself and checking the candidate's identification. It then continues as an interview.]

- What is the climate like in your country?
- Are there any problems with the climate in your country?
- What is your favourite type of weather?
- What kind of weather do you dislike?
- Would you prefer to live in a hot or cold country?
- What is the best/worst weather for travelling in?

Part 2 – Individual long turn (3 - 4 minutes)

You will have to talk about the topic for 1 to 2 minutes. You have one minute to think about what you're going to say. You can make some notes to help you if you wish.



Part 3 – Two-way discussion (3-4 minutes)

1. Which environmental problems are people most concerned about in your country?
2. What are the main causes of these environmental problems?

3. Do you think it is the responsibility of governments alone to protect the environment?
4. What measures can individuals take to protect the environment?
5. Do you think large companies and business organizations should be more environmentally friendly? Why? How?
6. How can we teach children about the importance of protecting the environment?

Retrieved from <http://www.ipassielts.com>.

Appendix B

Invitation to Test-takers



Are you a **Chinese-as-a-first-language graduate student**?

I am conducting a study on ratings assigned to EAL (English-as-an-Additional-Language) test-takers during the **IELTS speaking test**. The study will take place at the linguistics lab at the University of Victoria. You will have the opportunity to practice for the IELTS Speaking Test with a certified examiner and also receive your score at the end of the test.

My name is Christiani Thompson Wagner, and I am a graduate student in the department of Linguistics at the University of Victoria. You may contact me by e-mail (christianit2014@gmail.com) or phone (250) 472-5444 if you have questions.

The results of the study will:

1. Provide a current assessment of your English language oral skills
2. Help current and future IELTS test-takers by analyzing how ratings are assigned during the test
3. Help test-takers understand what elements are assessed during the IELTS speaking test

If you are interested in taking part of this study, please contact us by email (christianit2014@gmail.com). Participants will receive a \$10 Starbucks gift card for their participation.

Questions? Please email christianit2014@gmail.com.

Appendix C

Invitation to Examiners



My name is Christiani Thompson Wagner, and I am a graduate student in the department of Linguistics at the University of Victoria. You may contact me by e-mail (christianit2014@gmail.com) or phone (250) 472-5444 if you have questions.

I would like to invite English-as-a-first-language IELTS examiners to take part in a study I am conducting on ratings assigned to English-as-an-additional-language (EAL) test-takers during the IELTS speaking test. The study will take place at the linguistics lab at the University of Victoria.

If you are interested in being part of this study and live in Vancouver or on Vancouver Island, please contact me by email (christianit2014@gmail.com).

Questions? Please email christianit2014@gmail.com.

Appendix D

Examiners' Background Questionnaire

Part 1

| | | | |
|---|---------------------------------|--------------------------------|--------------------------------|
| 1. Letter you have been assigned on this research | A | | |
| 2. What is your age? | <input type="checkbox"/> 18-24 | <input type="checkbox"/> 25-34 | <input type="checkbox"/> 35-44 |
| | <input type="checkbox"/> 45-54 | <input type="checkbox"/> 55-64 | <input type="checkbox"/> 65-74 |
| 3. What is your gender? | <input type="checkbox"/> Female | <input type="checkbox"/> Male | <input type="checkbox"/> Other |

Part 2

| | | | |
|---|---|--------------------------------------|--------------------------------------|
| 4. Where were you born? | | | |
| 5. How long did you live in the country where you were born? | <input type="checkbox"/> Less than 1 year | <input type="checkbox"/> 1-5 years | <input type="checkbox"/> 6-10 years |
| | <input type="checkbox"/> 11-15 years | <input type="checkbox"/> 16-20 years | <input type="checkbox"/> 21-25 years |
| | <input type="checkbox"/> More than 25 years | | |
| | | | |
| 6. How many years have you spent in non-English speaking countries? | <input type="checkbox"/> Less than 1 year | <input type="checkbox"/> 1-5 years | <input type="checkbox"/> 6-10 years |
| | <input type="checkbox"/> 11-15 years | <input type="checkbox"/> 16-20 years | <input type="checkbox"/> 21-25 years |
| | <input type="checkbox"/> More than 25 years | | |
| | | | |

Part 3

| | |
|---|--|
| 7. In which countries have you taught English? | |
| 8. How many hours a day do you speak English with English language learners? | <input type="checkbox"/> Less than 1 <input type="checkbox"/> 1-2 hours <input type="checkbox"/> 3-4 hours <input type="checkbox"/> 5-6 hours <input type="checkbox"/> 7-8 hours <input type="checkbox"/> 9-10 hours <input type="checkbox"/> More than 10 hours |
| 9. For how long have you been an examiner? | <input type="checkbox"/> 1 year or less <input type="checkbox"/> 2-4 years <input type="checkbox"/> 4-6 years <input type="checkbox"/> 6-10 years <input type="checkbox"/> 10-12 years <input type="checkbox"/> over 12 years |
| 10. In what country(ies) have you worked as a speaking test examiner? | |
| 11. As an examiner, what are your thoughts about Chinese-as-a-first-language test-takers? | |

Appendix E

Test-takers' Background Questionnaire

Part 1

| | |
|---|--|
| 1. What number has been assigned to you on this research? | |
| 2. What is your age? | |
| 3. What is your gender? | |

Part 2

| | |
|--|--|
| 4. Where were you born? | |
| 5. How long did you live in the country where you were born? | |
| 6. At what age did you start studying English? | |
| 7. For how long did you study English? | |
| 8. How many years have you spent in English-speaking countries? | |
| 9. How many hours a day do you speak English with fluent speakers? | |

Part 3

| | |
|--|--|
| 10. How long have you been in Canada? | |
| 11. In how many English speaking tests have you participated? | |
| 12. Have you ever received training for the IELTS speaking test? | |
| 13. How many times have you taken the IELTS test? | |
| 14. What was your most recent score for the speaking test? | |

Appendix F

Modified Rating Sheet - Face-to-face Speaking Test

This rating sheet is intended to be used as a tool to assist examiners with feedback during the stimulated video recall. It provides an organized method of reviewing and rating the test-takers' performance during the speaking test.

| Please circle | | | | | | |
|---------------|---|---|---|---|---|---|
| EXAMINER | A | B | C | D | E | F |
| TEST-TAKER | 1 | 2 | 3 | 4 | 5 | 6 |

| Test-taker # | Fluency and coherence | Lexical resource | Grammatical range and accuracy | Pronunciation | Score |
|------------------|-----------------------|------------------|--------------------------------|---------------|-------|
| Test-taker _____ | | | | | |

Please comment on how you assigned the scores to this test-taker.

Appendix G

Band Descriptors

| Band | Fluency and coherence | Lexical resource | Grammatical range and accuracy | Pronunciation |
|------|--|--|---|---|
| 9 | <ul style="list-style-type: none"> speaks fluently with only occasional repetition or self-correction; any hesitation is content-related rather than to find words or grammar speaks coherently with fully appropriate cohesive features develops topics fully and appropriately | <ul style="list-style-type: none"> uses vocabulary with full flexibility and precision in all topics uses domestic language naturally and accurately | <ul style="list-style-type: none"> uses a full range of structures naturally and appropriately produces consistently accurate structures apart from 'slip' characteristics of native speaker speech | <ul style="list-style-type: none"> uses a full range of pronunciation features with precision and fluency sustains flexible use of features throughout is effortless to understand |
| 8 | <ul style="list-style-type: none"> speaks fluently with only occasional repetition or self-correction; hesitation is usually content-related and only rarely to search for language develops topics coherently and appropriately | <ul style="list-style-type: none"> uses a wide vocabulary resource readily and flexibly to convey precise meaning uses less common and idiomatic vocabulary skillfully, with occasional inaccuracies uses paraphrase effectively as required | <ul style="list-style-type: none"> uses a wide range of structures flexibly produces a majority of error-free sentences with only very occasional inaccuracies or less-than-systematic errors | <ul style="list-style-type: none"> uses a wide range of pronunciation features sustains flexible use of features, with only occasional lapses is easy to understand throughout, L1 accent has minimal effect on intelligibility |
| 7 | <ul style="list-style-type: none"> speaks at length without noticeable effort or loss of coherence may demonstrate language-related hesitation at times, or some repetition and/or self-correction uses a range of connectives and discourse markers with some flexibility | <ul style="list-style-type: none"> uses vocabulary resource flexibly to discuss a variety of topics uses some less common and idiomatic vocabulary and shows some awareness of style and collocation, with some inappropriate choices uses paraphrase effectively | <ul style="list-style-type: none"> uses a range of complex structures with some flexibility frequently produces error-free sentences, though some grammatical mistakes persist | <ul style="list-style-type: none"> shows all the positive features of Band 6 and some, but not all, of the positive features of Band 8 |
| 6 | <ul style="list-style-type: none"> is willing to speak at length, though may lose coherence at times due to occasional repetition, self-correction or hesitation uses a range of connectives and discourse markers but not always appropriately | <ul style="list-style-type: none"> has a wide enough vocabulary to discuss topics at length and make meaning clear in spite of inaccuracies generally paraphrases successfully | <ul style="list-style-type: none"> uses a mix of simple and complex structures, but with limited flexibility may make frequent mistakes with complex structures though these rarely cause comprehension problems | <ul style="list-style-type: none"> uses a range of pronunciation features with mixed control shows some effective use of features but this is not sustained can generally be understood throughout, though mispronunciation of individual words or sounds reduces clarity at times |
| 5 | <ul style="list-style-type: none"> usually maintains flow of speech but uses repetition, self-correction and/or slow speech to keep going may over-use certain connectives and discourse markers produces simple speech fluently, but more complex communication causes fluency problems | <ul style="list-style-type: none"> manages to talk about familiar and unfamiliar topics but uses vocabulary with limited flexibility attempts to use paraphrase but with mixed success | <ul style="list-style-type: none"> produces basic sentence forms with reasonable accuracy uses a limited range of more complex structures, but these usually contain errors and may cause some comprehension problems | <ul style="list-style-type: none"> shows all the positive features of Band 4 and some, but not all, of the positive features of Band 6 |
| 4 | <ul style="list-style-type: none"> cannot respond without noticeable pauses and may speak slowly, with frequent repetition and self-correction links basic sentences but with repetitive use of simple connectives and some breakdowns in coherence | <ul style="list-style-type: none"> is able to talk about familiar topics but can only convey basic meaning on unfamiliar topics and makes frequent errors in word choice rarely attempts paraphrase | <ul style="list-style-type: none"> produces basic sentence forms and some correct simple sentences but subordinate structures are rare errors are frequent and may lead to misunderstanding | <ul style="list-style-type: none"> uses a limited range of pronunciation features attempts to control features but lapses are frequent mispronunciations are frequent and cause some difficulty for the listener |
| 3 | <ul style="list-style-type: none"> speaks with long pauses has limited ability to link simple sentences gives only simple responses and is frequently unable to convey basic message | <ul style="list-style-type: none"> uses simple vocabulary to convey personal information has sufficient vocabulary for less familiar topics | <ul style="list-style-type: none"> attempts basic sentence forms but with limited success, or relies on apparently memorised utterances makes numerous errors except in memorised expressions | <ul style="list-style-type: none"> shows some of the features of Band 2 and some, but not all, of the positive features of Band 4 |
| 2 | <ul style="list-style-type: none"> cannot lengthily deliver more than a few words little communication possible | <ul style="list-style-type: none"> only produces isolated words or memorised utterances | <ul style="list-style-type: none"> cannot produce basic sentence forms | <ul style="list-style-type: none"> Speech is often unintelligible |
| 1 | <ul style="list-style-type: none"> no communication possible no intelligible language | | | |
| 0 | <ul style="list-style-type: none"> does not attend | | | |

Appendix H

Consent Form - Examiners



UNIVERSITY OF VICTORIA
DEPARTMENT OF LINGUISTICS

Participant Consent Form

A Preliminary Study of Ratings Assigned by English-as-a-First-Language Examiners to Chinese-as-a-first-language Test-takers during Simulated IELTS Speaking Tests

You are invited to participate in a study entitled *A Preliminary Study of Ratings Assigned by English-as-a-First-Language Examiners to Chinese-as-a-first-language Test-takers during Simulated IELTS Speaking Tests*.

My name is Christiani Thompson Wagner, and I am a graduate student in the department of Linguistics at the University of Victoria. You may contact me by e-mail (christianit2014@gmail.com) or phone (250) 472-5444 if you have questions.

As a graduate student, I am required to conduct research as part of the requirements for a degree in Master's in Applied Linguistics. This study is being conducted under the supervision of Dr. Li-Shih Huang. You may contact my supervisor at lshuang@uvic.ca or at (250) 472-4665.

Purpose and Objectives

The purpose of this research project is to examine elements of communication involved in ratings assigned by English-as-a-first-language IELTS examiners to Chinese-as-a-first-language test-takers during the IELTS speaking test.

Importance of this Research

Research of this type is important because it may increase examiners' awareness of elements that affect rates assigned during language proficiency tests.

Participants Selection

You are being asked to participate in this study because you are an English-as-a-first-language examiner with at least 10 years of experience in oral language assessment.

What is involved:

If you consent to voluntarily participate in this research, your participation will include performing and rating one simulated IELTS speaking test with a Chinese-as-a-first-language graduate student as well as rating 6 audio-recorded simulated IELTS speaking tests done by other participants.

The study will take two hours and thirty minutes. You will be required to:

- Sign a consent form – 5 minutes
- Fill out a questionnaire – 5 minutes

- Administer a speaking test – 11 to 14 minutes
- Participate in a video-stimulated verbal recall session – 30 minutes
- Rate 6 audio-recorded simulated IELTS speaking tests, which will be emailed to you – 90 minutes

Data will be collected through video and audio recording and one questionnaire.

Video recording and audio recording will be made, and, written notes will be taken. A transcription of the speaking test and of the video stimulated verbal recall session will be made.

Inconvenience

Participation in this study may cause some inconvenience to you, such as the time you will spend traveling to the linguistics lab at the University of Victoria and taking part in the study, and stress or fatigue. If you feel stressed or tired during the study, you will have the option of taking a mini break, which may include leaving the room for a few minutes, going to the washroom, getting a drink of water, etc.

Risks

There are no known or anticipated risks to you by participating in this research.

Benefits

The potential benefits of your participation in this research include a better understanding of the IELTS speaking test and of how ratings are assigned to test-takers.

Compensation

You will be paid \$25/hour.

Voluntary Participation

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation. If you do withdraw from the study, your data will only be used if you give permission.

If you decide to withdraw during the data collection session or anytime thereafter, incentives will be given and will be pro-rated.

On-going Consent

To make sure that you continue to consent to participate in this research, I will remind you that your participation is voluntary and that, if you decide to participate, you may withdraw at any time without any consequences or explanation. I will ask you to provide a verbal consent, which will be documented by audio recording. This will take place before the second session, when you will rate audio-recorded speaking tests.

Anonymity

In terms of protecting your anonymity, all examiners will be assigned letters. All data will be coded.

The small size of the sample may, however, make it possible for the examiner and the test-taker to identify each other. It is also possible that participants may be identified through participation.

Confidentiality

Your confidentiality and the confidentiality of the data will be protected. All data will be password-protected, encrypted computer files that will be stored in an IronKey USB flash drive.

All data (video and audio recordings, questionnaires, transcripts, rating sheets, and researcher notes) will be kept in my office (Clearihue D271 at the University of Victoria) in a locked filing cabinet.

Video and audio tapes will not be shared with the public and will be used only for analysis.

Dissemination of Results

It is anticipated that the results of this study will be shared with others in the following ways:

- Directly to participants

- Thesis dissertation and presentation
- Published article
- Presentation at scholarly meetings

Disposal of Data

Data from this study will be disposed. Computer files will be permanently destroyed, and paper documents will be shredded five years upon completion of the study. All video and audio files will be permanently deleted off camera's hard drive immediately after loaded onto IronKey USB flash drive. Video and audio files will also be permanently deleted five years upon completion of the study.

Contacts

Individuals that may be contacted regarding this study include the researcher, Christiani P. Thompson Wagner, and the researcher's supervisor, Dr. Li-Shih Huang. In addition, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the Human Research Ethics Office at the University of Victoria (250-472-4545 or ethics@uvic.ca).

Your signature below indicates that you understand the above conditions of participation in this study, that you have had the opportunity to have your questions answered by the researchers, and that you consent to participate in this research project.

Name of Participant

Signature

Date

A copy of this consent will be left with you, and a copy will be taken by the researcher.

Appendix I

Consent Form - Test-takers



UNIVERSITY OF VICTORIA
DEPARTMENT OF LINGUISTICS

Participant Consent Form

A Preliminary Study of Ratings Assigned by English-as-a-First-Language Examiners to Chinese-as-a-first-language Test-takers during Simulated IELTS Speaking Tests

You are invited to participate in a study entitled *A Preliminary Study of Ratings Assigned by English-as-a-First-Language Examiners to Chinese-as-a-first-language Test-takers during Simulated IELTS Speaking Tests*.

My name is Christiani Thompson Wagner, and I am a graduate student in the department of Linguistics at the University of Victoria. You may contact me by e-mail (christianit2014@gmail.com) or phone (250) 472-5444 if you have any questions.

As a graduate student, I am required to conduct research as part of the requirements for a degree in Master's in Applied Linguistics. This study is being conducted under the supervision of Dr. Li-Shih Huang. You may contact my supervisor at lshuang@uvic.ca or at (250) 472-4665.

Purpose and Objectives

The purpose of this research project is to examine elements of communication involved in ratings assigned by English-as-a-first-language examiners to Chinese-as-a-first-language test-takers during the IELTS speaking test.

Importance of this Research

Research of this type is important because it may increase international students' awareness of elements that affect rates assigned during language proficiency tests.

Participants Selection

You are being asked to participate in this study because you are a Chinese-as-a-first-language graduate student.

What is involved:

If you consent to voluntarily participate in this research, your participation will consist of taking part in a simulated IELTS speaking test with an English-as-a-first-language examiner.

The study will take 30 minutes. You will be required to:

- Sign consent form – 5 minutes
- Fill out a background questionnaire – 5 minutes
- Perform a speaking test – 11 to 14 minutes

Data will be collected through video and audio recording and one questionnaire.

Video recording, audio recording, written notes, and observations will be made. A transcription of the video stimulated verbal recall session will be made.

Inconvenience

Participation in this study may cause some inconvenience to you, such as the time you will spend traveling to the linguistics lab at the University of Victoria and taking part in the study, and stress even though this is not a real speaking test. The scores will not affect your academic records or future IELTS records. If you feel stressed during the speaking test, you will have the option of taking a mini break, which may include leaving the room for a few minutes, going to the washroom, getting a drink of water, etc.

Risks

As this study involves simulated IELTS speaking tests, you will not be able to use the test scores that are assigned to you.

Benefits

The potential benefits of your participation in this research include practicing for the IELTS speaking test for free with an examiner and receiving your score upon completion of the test.

Compensation

As a way to compensate you for your time, you will be given a \$10 gift card. If you consent to participate in this study, this form of compensation to you must not be coercive. It is unethical to provide undue compensation or inducements to research participants. If you would not participate if the compensation were not offered, then you should decline.

Voluntary Participation

Your participation in this research must be completely voluntary. If you do decide to participate, you may withdraw at any time without any consequences or any explanation. If you do withdraw from the study your data will only be used if you give permission.

If you decide to withdraw during the data collection session or anytime thereafter, incentives will be given.

Anonymity

In terms of protecting your anonymity, all test-takers will be assigned numbers. All data will be coded. The small size of the sample may, however, make it possible for the examiner and the test-taker to identify each other. It is also possible that participants may be identified through participation.

Confidentiality

Your confidentiality and the confidentiality of the data will be protected. All data will be password-protected, encrypted computer files that will be stored in an IronKey USB flash drive.

All data (video and audio recordings, questionnaires, transcripts, rating sheets, and researcher notes) will be kept in my office (Clearihue D271 at the University of Victoria) in a locked filing cabinet.

Video and audio tapes will not be shared with the public and will be used only for analysis.

Dissemination of Results

It is anticipated that the results of this study will be shared with others in the following ways:

7. Directly to participants
8. Thesis dissertation and presentation
9. Published article
10. Presentation at scholarly meetings

Disposal of Data

Data from this study will be disposed. Computer files will be permanently destroyed, and paper documents will be shredded five years upon completion of the study. All video and audio files will be permanently deleted off camera's hard drive immediately after loaded onto IronKey USB flash drive. Video and audio files will also be permanently deleted five years upon completion of the study.

Contacts

Individuals that may be contacted regarding this study include the researcher, Christiani P. Thompson Wagner, and the researcher's supervisor, Dr. Li-Shih Huang. In addition, you may verify the ethical approval of this study, or raise any concerns you might have, by contacting the Human Research Ethics Office at the University of Victoria (250-472-4545 or ethics@uvic.ca).

Your signature below indicates that you understand the above conditions of participation in this study, that you have had the opportunity to have your questions answered by the researchers, and that you consent to participate in this research project.

A copy of this consent will be left with you, and a copy will be taken by the researcher.

Appendix J

Audio-recorded Speaking Test - Rating Sheet

| Please circle | | | | | | |
|---------------|---|---|---|---|---|---|
| EXAMINER | A | B | C | D | E | F |

| Test-taker | Fluency and coherence | Lexical resource | Grammatical range and accuracy | Pronunciation | Rating |
|------------------|-----------------------|------------------|--------------------------------|---------------|--------|
| Test-taker _____ | | | | | |
| Test-taker _____ | | | | | |
| Test-taker _____ | | | | | |
| Test-taker _____ | | | | | |
| Test-taker _____ | | | | | |

Please comment on how you assigned scores to this test-taker:

Appendix K

Coding Scheme

| Nonverbal behaviour | Test-taker # | Time (minutes & seconds) |
|---|--------------|--------------------------|
| Eye contact | | |
| Looks to the left/right (away from examiner) | | |
| Looks up/down (away from examiner) | | |
| Looks to the left/right (up) away from examiner | | |
| Looks to the left/right (down) away from examiner | | |
| Smiles | | |
| Moves left/right hand to the left | | |
| Moves left/right hand to the right | | |
| Brings right/left/ both hands close to the chest | | |
| Moves right/left/both hands away from the chest | | |
| Brings hands together (palms and/or fingers together) | | |
| Touches <u>object</u> (description of movement) | | |
| Touches <u>self</u> (description of movement) | | |
| Extends right/left arm out | | |
| Brings left/right arm in | | |
| Moves left/right/both hands hand up | | |
| Moves left/right/both hands hand down | | |
| Holds <u>object</u> with right/left/both hands | | |
| Grips _____ with left/right/both hands | | |
| Turns left/right/both hands up/down | | |
| Closes fingers of right/left/both hands | | |
| Moves left/right/both hands in circular motion | | |
| Opens left/right/both hands | | |

| | | |
|--|--|--|
| Closes left/right/both hands | | |
| Brings back of hands together | | |
| Brings fingers of left/right hand together | | |
| Covers left/right hand with right/left hand | | |
| Moves left/right <u>finger</u> /thumb up/down | | |
| Moves left/right/both hands back-and-forth to the left/to the right (number of times) | | |
| Makes _____ shape with left/right/both hands | | |
| Points to _____ with the left/right/both hands | | |
| Further detailing of each movement, inclusion of less frequent movements, and inclusion of parts of the speech that accompanied the nonverbal cue were added when I believed the information could help me conduct a more detailed analysis of the data. | | |

Nonverbal Cues by Type

| Nonverbal cue | Sub-category | Environment | Examiner | Test-taker |
|---------------|-----------------|------------------|----------|------------|
| Smiles | Duchenne | In Isolation | | |
| | | Alongside Speech | | |
| | Non-Duchenne | In Isolation | | |
| | | Alongside Speech | | |
| Gestures | Illustrators | In Isolation | | |
| | | Alongside Speech | | |
| | Self-adaptors | In Isolation | | |
| | | Alongside Speech | | |
| | Object-adaptors | In Isolation | | |
| | | Alongside Speech | | |
| Eye contact | | In Isolation | | |
| | | Alongside Speech | | |