INTERLOCUTOR BEHAVIORS DURING ORAL ASSESSMENT INTERVIEWS

A Master's Thesis

by

Gülay Akın Şahin

The Department of

Teaching English as a Foreign Language

Bilkent University

Ankara

July 2010

INTERLOCUTOR BEHAVIORS DURING ORAL ASSESSMENT INTERVIEWS

The Graduate School of Education

of

Bilkent University

by

GÜLAY AKIN ŞAHİN

In Partial Fulfillment of the Requirements for the Degree of
Master of Arts
in

The Department of

Teaching English as a Foreign Language

Bilkent University

Ankara

July 2010

BILKENT UNIVERSITY

THE GRADUATE SCHOOL OF EDUCATION

MA THESIS EXAMINATION RESULT FORM

July 9, 2010

The examining committee appointed by The Graduate School of Education for the thesis examination of the MA TEFL student

Gülay Akın Şahin

has read the thesis of the student.

The committee has decided that the thesis of the student is satisfactory.

Thesis title: Interlocutor Behaviors during Oral Assessment Interviews

Thesis Advisor: Prof. Dr. Kimberly Trimble
Bilkent University, MA TEFL Program

Committee Members: Asst. Prof. Dr. Philip Durrant
Bilkent University, MA TEFL Program

Dr. Deniz Şallı-Çopur
Middle East Technical University,
Department of Foreign Language Education

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Teaching English as a Second Language.

_____

(Prof. Dr. Kimberly Trimble)
Supervisor

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Teaching English as a Second Language.

_____

(Asst. Prof. Dr. Philip Durrant)
Examining Committee Member

I certify that I have read this thesis and have found that it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Teaching English as a Second Language.

_____

(Dr. Deniz Şallı-Çopur)
Examining Committee Member

Approval of the Graduate School of Education

_____

(Visiting Prof. Dr. Margaret Sands)
Directo

ABSTRACT

INTERLOCUTOR BEHAVIORS DURING ORAL ASSESSMENT INTERVIEWS

Gülay Akın Şahin

Master of Arts in Teaching English as a Foreign Language

Supervisor: Prof. Dr. Kimberly Trimble

July 2010

This study investigated interlocutor behaviors during oral assessment interviews. The study was conducted at Karadeniz Teknik University, with the participation of 272 students and twenty instructors assigned to conduct oral interviews in the Department of Basic English.

Data were collected through questionnaires and the analysis of video-recordings. Students ranked each of the interviewers separately. The analysis of the quantitative data revealed that while students generally perceived their interlocutors' behaviors more positively, there were important differences among the behaviors. The analysis of the video-recordings also showed that the interlocutors generally displayed positive behaviors during oral assessment interviews but showed variation in their behaviors.

Key words: Oral assessment interview, interlocutor behavior

ÖZET

SÖZLÜ MÜLAKAT SINAVLARINDA KONUŞMACININ DAVRANIŞLARI

Gülay Akın Şahin

Yüksek lisans, Yabancı Dil Olarak İngilizce Öğretimi Bölümü

Tez Yöneticisi: Prof. Dr. Kimberly Trimble

July 2010

Bu çalışma sözlü mülakat sınavlarındaki konuşmacı hocaların davranışlarını ortaya çıkarmayı amaçlamıştır. Çalışma, Karadeniz Teknik Üniversitesi, Temel İngilizce Bölümündeki 272 öğrenci ve mülakat sınavlarında konuşmacı olarak görevlendirilen yirmi öğretim elemanının katılımıyla gerçekleştirilmiştir.

Veriler, anket ve video kayıtlarının analiziyle oluşturulmuştur. Öğrenciler her iki konuşmacıyı ayrı ayrı değerlendirdiler. Nicel veri analizinin sonuçları, öğrencilerin, mülakatlarındaki konuşmacıların davranışlarını daha çok pozitif olarak algıladıklarını ancak davranışlar arasında önemli farklılıklar olduğunu ortaya çıkarmıştır. Video kayıtlarının analizi de konuşmacıların, sözlü mülakat sınavı boyunca genellikle pozitif davranışlar sergilediğini fakat davranışlarında farklılıklaştıklarını göstermiştir.

Anahtar Sözcükler: Sözlü mülakat sınavı, konuşmacı davranışı

ACKNOWLEDGEMENTS

Writing a thesis was an enjoyable but demanding journey for me. Throughout this journey, I had a number of people beside me whose encouragement and assistance I would not ignore. First and foremost, I would like to express my thanks and deepest gratitude to my thesis advisor, Prof. Dr. Kimberly Trimble, for his guidance, support and patience. Without his invaluable contributions and perfectionism, this thesis would have never been completed.

I would like to extend my appreciation to Asst. Prof. Dr. Julie Matthews Aydınlı, whose smiling face, endless positive look at the things helped me to survive in this demanding program. I also owe much to Asst. Prof. Dr. Philip Durrant for his guidance, whole-hearted support and patience. Whenever I had a question, he was there to help me.

I would also like to thank to Asst. Prof. Dr. Naci Kayaoğlu, the director of Karadeniz Teknik University, School of Foreign Languages, who gave me the permission to attend the MA TEFL program.

My special thanks go to my friend A. Ulus Kımav and Hawraz Q. Hama for his help and support.

I am indebted too much to my friends Onur Dilek, Işıl Dilek and Yonca Özsandıkçı whose help and support always kept me strong and persistent in my efforts during this challenging year.

Last but not least, I would like to express my deep appreciation to my dear husband Fatih Şahin and my family for their endless love, encouragement and tolerance and always being beside me for all my life not necessarily for this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

Introduction

Assessment has been one of the core issues of the teaching and learning process with many issues still unresolved. Despite these on-going discussions, the role of the assessment procedure, which has been agreed on by many researchers, has been defined as not only providing the teachers and students with feedback in order to help them shape future goals for learning, but also informing both students and teachers about successes or deficiencies in a particular learning process. In other words, assessment is fundamental; it is a way of understanding to what extent the aims of teaching have been accomplished.

Although assessment is regarded as something to enhance or improve the teaching and learning processes for students and teachers, many students feel stressed, anxious, and inadequate when they have any kind of assessment. This may be in part because our assessment, as teachers, does not feed back into students' learning but only serves for grading or marking, which results in great pressure on students to succeed. The feeling of anxiety can increase when it comes to students' oral assessment in the form of one-on-one interview with the assessor. There are many studies which have been conducted to find out the possible causes of the anxiety (Brown, 2003; Lazaraton, 1996; McNamara, 1997; Oya, Manalo, & Greenwood, 2004). However, the interlocutor's behavior during a teacher-student interview has been little focused on. Additionally, the variation in the behaviors has not been studied for students with beginner level students.

With this aim, this study examined the interlocutors' positive and negative behaviors during oral performance assessment. In the study, both students' affective responses to their interlocutors as well as observations of the interlocutors during interviews were used to investigate these behaviors.

Background of the study

While widely used, oral proficiency interviews have been considered as a highly controversial type of assessment in speaking proficiency tests. As O'Sullivan (2000) mentioned, performance in language test tasks can be influenced by a wide range of features, which can interact unpredictably with characteristics of individual test-takers (cited in Lumley & O'Sullivan, 2005). Although there has been much research which highlights the factors that affect the candidates' output during the assessment procedure, this research remains inconclusive, with a range of variables identified by a variety of studies.

Oral interviews are mainly conducted in three formats: individually, in pairs, and in groups. Among the reasons that have been identified as influencing differences in the candidates' performance in each type of interview tests, the effects of the interlocutor behaviors has been the focus of many researchers. In addition to a number of variables associated with interlocutors, such as age, sex, proficiency and personality, Lumley and O'Sullivan (2005) studied the effects of task topic, the gender of the person presenting the topic and the gender of the candidate. They conducted their study in a tape-based speaking test, which was delivered in a language laboratory, with no interlocutor present, but where stimulus material was presented by one or more speakers, one of whom acted as audience for the candidate's speech. The study found that there were small effects for some of the

hypothesized interactions. When the students were required to talk about a topic they were unfamiliar with to a hypothetical foreign male, the students were more threatened than when they showed their ignorance about the topic to an absent female.

The influence of interlocutor proficiency in a paired oral assessment has also been considered as an important issue in the assessment process. In his study concerning this issue, Davis (2009) examined a group of twenty first-year students at a Chinese university. The students were grouped with partners who were of relatively high and low proficient in English. Although the results of the study showed that the proficiency level of the interlocutor appeared to have little effect on average raw scores of the participants, it was observed that examinees with lower-level proficiency produced more language words when they worked with a higher-level partner. Moreover, the students with lower-levels had more confidence in themselves when grouped with higher-level students.

Looking at the interlocutor effect in individual interviews, in which there is a test-taker and a teacher, we see that there are many behaviors of interlocutors that are considered to affect the candidates' performance. Among these behaviors, Lumley and Brown (1996) have discovered that a number of features of the interlocutor behavior appear to affect the level of difficulty of the interaction for the candidate (cited in McNamara, 1997). In their study, they examined the assessments of seventy offshore candidates from two administrations of the Occupational English Test. The recordings of the interviews were rated for perceptions of the competence of the interlocutor and the rapport established between the candidate and the interlocutor. While some of the interlocutors helped the students by scaffolding and creating a

supportive atmosphere, some of them did not, which changed the candidates' final scores. McNamara & Lumley (1997) point out that the candidate's score is clearly the outcome of the interaction of variables, one of which is the interlocutor's behavior.

Similarly, concerning the effect of the variations in the interlocutors' behaviors on candidates' performance, Brown (2003) states that although interviewers are generally provided with guidelines that suggest topics and general questioning focus, when specific questions are neither preformulated nor identical for each candidate, this may lead to changes in the assessment of speaking ability of the candidates. Brown (2003) also points out that the attitudes of the interlocutors towards candidates may vary, which may affect the candidates' performance.

In reviewing the literature, Brown (2003) identified studies which focused on the differences in aspects of the interlocutors' behaviors, such as the level of rapport that they establish with the candidates (Lazaraton, 1996; T. F. McNamara, 1997); their functional and topical choices (Brown & Lumley, 1997; Reed & Halleck, 1997); the ways in which they ask questions and construct prompts (Brown & Lumley, 1997; Lazaraton, 1996); and the ways in which they accommodate their speech to the speech of the candidate (Brown & Lumley, 1997; Lazaraton, 1996; Ross, 1992; Ross & Berwick, 1992); and the ways in which they develop and extend topics (Berwick & Ross, 1996). Despite this large body of research, however, how interlocutors support of or hinder to the candidates linguistically and interactionally remain unclear.

Another point that is thought to be relevant to interlocutor behavior is the opportunities given by the interlocutor to the candidate to demonstrate his

competence during the assessment process. Lazaraton (1996) explains that we cannot ensure that all candidates are given the same number and kinds of opportunities to display their abilities unless examiners conduct themselves in similar, prescribed ways. Moreover, Brown (2003) found that interviewers could make the candidate appear to be an effective communicator by their scaffolding, explicit questioning, smoothly extending topics and providing frequent positive feedback. On the other hand, they could make the candidate appear to be a poor communicator by confusing the candidates with frequent topic shifts and using ambiguous closed questions to elicit extended answers (as cited in Fumiyo Nakatsuhara, 2008).

In order to scrutinize the discourse between the teacher and the student talk during the assessment talk from a different perspective, Nakamura (2008) observed what happens when a teacher talks to a student outside of the classroom setting. He described features of 'repair' as they occur in a sequence of turns to describe how this particular discourse genre is co-accomplished. He revealed that once the purpose of the talk moves beyond controlled production of correct language forms, the focus of most oral interviews, the interlocutors' roles and relationship shift from expert and novice to co-participants in managing the talk. Nakamura (2008, p. 272) summarizes the ideal nature of this interaction underlining that ''the roles and purposes are dynamically interwoven into the fabric of talk-interaction."

The literature on assessment contains a number of studies about factors which are considered to have a positive effect on candidates output in oral assessments. Although, many studies, such as O'Sullivan (2002), Brown (2003), Davis (2009), Ockey (2009), reached some conclusions about the effects of interlocutors, these effects cannot be generalized to Turkish EFL context. Additionally, these variations

in the interlocutors' behaviors need to be analyzed with beginner level students, where students may have less control over the interview process because of language competency levels, to determine candidates are treated in the same way and to explore the students' perceptions of their interlocutor behaviors.

With this aim, in this study, students' perceptions and observations of interviews will be used to examine interlocutors' positive and negative behaviors during oral performance assessment.

Statement of the Problem

Oral interviews, in which examiner and candidate take part in an unscripted discussion of general topics, have long been a popular method for assessing oral second language proficiency (Brown, 2003) . A great deal of research has been conducted on the strengths and weaknesses of this type of testing system. Studies about oral assessment interviews have varied widely on the effect of test-taker gender, audience and topic on task performance (Lumley & O'Sullivan, 2005). Other research has looked at factors that affect the candidates' performances in a paired or group work test systems, including the effects of group members' personalities on a test taker's L2 group oral discussion test score (Ockey, 2009) and learners' familiarity with their partner in an interactive task.

Among those studies which looked specifically at the interlocutor's role in a teacher-student interview, Brown (2003) has examined the interviewers' variation in their ways of eliciting demonstrations of communicative ability, its impacts on the candidate's oral production, and the raters' perceptions of candidate ability. Another study conducted by Nakatsuhara (2008) has clearly exemplified possible relationship between the characteristics of interviewer behavior and particular components of

language ability affected. However, no research has been done to investigate the interlocutors' behaviors during a teacher-student interview with beginner level students and the candidates' perceptions of their interlocutors' behaviors. The purpose of this study is to examine what positive and negative behaviors the interlocutors display during oral performance assessment of beginner level students.

At Karadeniz Technical University, many of the teachers who are assigned to administer oral assessment tests complain that students are less successful during these interviews than they are during their speaking course. On the other hand, students complain that they are treated in a different way during their oral interviews by their interlocutors. The study looked at instructors' behaviors to determine if differences existed among these interviewer behaviors and if so, what these differences are.

Research Questions

This study will investigate the following research questions:

- What are students' perceptions of their interlocutor behaviors during oral assessment interviews?
- What are observed positive and negative behaviors of interlocutors during oral assessment interviews?

Significance of the Study

Because oral interviews play a major role in the assessment of speaking ability, their reliability is critical for judging the students' oral competence. One possible way for assuring the reliability is to understand the effect of the interlocutor during these interviews. However, the literature has little research on the effects of interlocutors on students at beginner levels or on students' perceptions of their

interlocutors' behaviors. Thus, this study may contribute to the literature by showing the variation in the interlocutors' behaviors during the oral performance assessment of beginner level students and their perceptions of these behaviors of interlocutors.

At the local level, the current oral assessment procedure, at my home institution of Karadeniz Technical University (KTU), aims to compensate for the students' concerns during the process of assessment and create as natural a conversational environment as possible. The results of this study may help those responsible for designing the oral assessments by contributing to the creation of guidelines to shape interviews in a way that allow students at various proficiency levels to demonstrate their full oral competence.

Conclusion

In this chapter, the background of the study, statement of the problem, research questions, and significance of the study have been presented. The next chapter will review the literature related to the purpose of the study. The third chapter will give detailed information about the methodology, including the setting, participants, instruments, and data collection and analysis procedures of the study. The fourth chapter will present the data analysis procedures and findings. Finally, the fifth chapter will present the discussion of the findings, limitations of the study, suggestions for further research, and pedagogical implications.

CHAPTER II: REVIEW OF LITERATURE

Introduction

This research study investigated the interlocutors' behaviors during oral performance tests. Students' perceptions as well as observations of video-recordings of interviews were gathered and analyzed. Based upon these analyses, recommendations were made to teachers who are assigned to conduct oral interviews to help them plan and conduct oral tests that enable students to demonstrate their real oral competence.

This chapter reviews the literature on testing speaking. The chapter consists of three sections. In the first section, the literature on assessment in general and performance-based assessment in particular will be briefly reviewed, including information on its strengths and limitations. The second section covers assessment of speaking skill, problems in testing speaking, and reliability and validity in speaking tests. The third section covers formats of testing speaking, paired and oral interviews, and the interlocutor effect on students in oral interviews.

Assessment

Assessment includes any means of checking of what students can or cannot do with the language (Abbott & Wingard, 1990). It may be seen "as a language, as a form of communication between the teacher and the pupil, the former seeking to identify both achievements and needs, the latter seeking to respond to signals of guidance as to appropriate subsequent activity" (Golstein & Lewis, 1996, p. 32). On the other hand, Goldstein and Lewis claim that assessment is, like any language, both delicate and far from perfect as a mechanism of communication. It is a form of

communication which entails a very high level of professional skill and understanding to articulate effectively (Golstein & Lewis, 1996). In an attempt to emphasize the importance of assessment, Goldstein and Lewis (1996) also stated that while there are three message systems in the classroom--curriculum, pedagogy and assessment-- it is assessment that provides the vital bridge between what is offered by the teacher and the reaction to it by the pupil (Bernstein, 1977, cited in Golstein & Lewis, 1996). That is, it is a process for obtaining information that is used for making decisions about students, curricula and programs.

Depending upon the purpose and target group, different types of assessment can be applied. These are norm-referenced assessment, criterion-referenced assessment, and performance –based assessment. Since the primary focus of this study is the assessment of speaking ability, we will focus on performance-based assessment type.

Performance-based Assessment

Performance-based assessment assesses what a student can do under specific but generalizable circumstance. Gipps (1994) provides a definition which specifies the main feature of performance-based assessment: “Performance measurement calls for examinees to demonstrate their capabilities directly, by creating some product or engaging in some activity” (p. 99). In addition, with respect to the content of the assessment process, Hinett & Thomas (1999) say that ‘‘A criterion must be supplied and be made available and transparent to the student; it is then assessed in terms of how well the performance matches the criterion specified” (p. 8).

The use of performance-based assessment has some advantages over traditional types of assessment. Baron & Wolf (1996) claim that there is growing evidence that

when compared with more traditional forms of assessment, performance-based assessment has a greater likelihood of increasing the capacity of teachers and students by providing models of learning and assessment events that are coherent with the best practices defined in a variety of disciplines.

As the emphasis in language classrooms have shifted from classical approaches to more communicative approaches, researchers have had to address the problem of how to measure the performance of students. They have had to "grapple with questions related to identifying what is important in their discipline, defining the characteristics of strong student work, setting standards for how much is good enough" (Baron & Wolf, 1996, p. 189).

Regarding the difference between the format of a performance-based assessment and the traditional assessment, McNamara (1996) presents two factors: "…a performance by the candidate which is observed and judged using an agreed judging process" (as cited in Karslı, 2002, p. 10).

Performance-based assessment has both strengths and weaknesses. Gipps (1994) states that performance-based assessment gives the examiner the ability to see whether students can use their knowledge communicatively, which is not possible in traditional ways of assessment. On the other hand, performance-based assessments may have some practical disadvantages, such as time and effort, since the students are assessed individually rather than in groups. In addition, the scoring may contain subjectivity to some degree, which decreases the reliability of the final result.

Assessment of Speaking Ability

Luoma (2004) emphasizes the importance of speaking ability saying that the ability to speak in a foreign language is at the core of what is means to be able to use

a foreign language. She also underlines a number of factors which contribute to spoken performance in a foreign language, such as our personality, self-image, knowledge of the world, and our ability to reason and express our thoughts. Speaking in a foreign language is very hard and it takes a long time to be competent since one must learn to **"**master the sound system of the language, have almost instant access to appropriate vocabulary and be able to put the words together intelligibly with minimal hesitation" (Luoma, 2004, p. 9).

Speaking is also the most challenging and difficult language skill to assess. A learner's speaking ability is usually assessed during a face-to-face interaction, in real time, between an interlocutor and a candidate (Luoma, 2004). Commenting on the reliability in this type of assessment, Luoma (2004) states that the assessor has to make instantaneous judgments about a range of aspects of what is being said, as it is being said. Hence, Harris (1969) claims that oral ratings have tendency to have rather low reliability. Harris also argues that there are no two interviews which are conducted exactly alike, even by the same interviewer.

Similarly, there is no interviewer who can maintain exactly the same scoring standards throughout a large number of interviews (Harris, 1969). The shifts both in the ways interviews are conducted and the interviewers themselves lower the reliability of final scoring. This means that ''the assessment might depend not only upon which particular features of speech (e.g. pronunciation, accuracy, fluency) the interlocutor pays attention to at any point in time, but upon a host of other factors such as the language level, gender, and status of the interlocutor, his or her familiarity to the candidate and the personal characteristics of the interlocutor and the candidate'' (Luoma, 2004, p. 10).

Moreover, the nature of the interaction, the kinds of task that are presented to the candidate, the topics discussed, and the opportunities provided to show his or her ability to speak in a foreign language may all have impact on the candidate's performance. From a testing perspective, Luoma (2004) considers speaking special because of its interactive nature. She states that the discourse during the assessment process cannot be predicted, just as no two conversations are alike even if they are about the same topic. Luoma (2004) points that special procedures are needed to ensure the reliability and validity of the scores.

Validity

Validity is considered as the most important factor in test development. "A test is said to be valid if it measures accurately what is intended to measure" (Hughes, 2003, p. 26). Although the term ''construct validity'' is used as a general notion of validity, there are also some subordinate forms of validity, including content validity and criterion-related validity. "A test is said to have content validity if its content constitutes a representative sample of the language skills, structures, etc. with which it is meant to be concerned" (Hughes, 2003, p. 26). Criterion-related validity "relates to the degree to which results on the test agree with those provided by some independent assessment of the candidate's ability"(Hughes, 2003, p. 27). There are two kinds of criterion-related validity: concurrent validity which is established when the test and the criterion are administered at about the same time and predictive validity which concerns the degree to which a test can predict candidate's future performance (Hughes, 2003).

Validity in scoring is also important in the assessment process. In order to be able to say that a test has validity, not only the items but also the responses scored

must be valid. There is no use having perfect items if they are not scored validly. Another subordinate form of validity is face validity. “A test is said to have face validity if it looks as if it measures what it is supposed to measure’’ (Hughes, 2003, p. 33). Hughes (2003) says that if a test aims to measure pronunciation ability but does not require the test taker to speak it may be thought to lack face validity.

Reliability

Luoma (2004) identified some studies in the literature which defined reliability as score consistency. Luoma (2004) states if the scores from a test given today are reliable, they will not change largely when the same test is given to the same people again. Reliability is important because it shows that the scores are dependable, so they can be relied on in making decision. On the other hand, unreliable scores “can lead to wrong placements, unjustified promotions, or undeservedly low grades on report cards” (Luoma, 2004, p. 176).

Formats of Speaking Tests

To assess the students’ oral ability, performance-based tests are generally used. The formats of testing speaking can be grouped under two headings: direct tests, which means that there is face to-face interaction with a human interlocutor such as interviews, role plays, or interpreting, and indirect tests in which there is no face-to-face interaction with an interlocutor (sometimes characterized as ‘semi-direct’ in some of the literature), such as prepared monologues and readings aloud. Hughes (2003) lists three general formats for speaking tests: Response to audio-or video recordings, interaction with fellow candidates, and interviews. The three formats and their advantages, and disadvantages are discussed below.

Responses to audio-or video recordings are often described as ‘semi-direct’ types of assessment. In this format, “candidates are presented with the same computer generated or audio-/video-recorded stimuli to which the candidates themselves respond in the microphone” (Hughes, 2003, p. 122). This type of semi-direct assessment is generally thought to have reliability. Moreover, it is practical and economical if a language laboratory is available, since a large number of students can be tested at the same time. On the other hand, the clear disadvantage of this format is “its inflexibility: there is no way of following up candidates’ responses” (Hughes, 2003, p. 122). Another disadvantage of this type of assessment is that some students may have trouble in using audio or video aids (Karslı, 2002).

In interaction with fellow candidates, two or more candidates may be asked to discuss a topic or make plans. This format enables students to be highly interactive during the assessment process. Hughes (2003) emphasizes the advantages of this type of assessment, saying that it should elicit language that is appropriate to exchanges between equals, which may well be called for in the test specifications. On the other hand, the problem with this type is that the performance of one candidate can be likely affected by that of the others. Similarly, if one of the candidates is more proficient, this may also affect their performance and accordingly their scores, too. Hence, if interaction with fellow candidates is to take place, Hughes (2003) suggests that the number of the candidates should not be more than two since with large numbers the chance of a shy candidate to show their ability decreases.

Interviews are the most common format for testing speaking. Madsen (1983) describes the nature of the interview saying that students are actually talking with someone instead of simply reciting information. In this sense, the oral interview can

provide a genuine sense of communication. Brown (2003) emphasizes the place of oral interviews in testing speaking and points that “oral interviews, in which examiner and candidate take part in an unscripted discussion of general topics, have long been a popular method for assessing oral second language proficiency” (p. 1). In 1996, Ross (cited in Brown, 2003) also states that since oral interviews entail features of nontest or conversational interaction, they help second language learners to demonstrate their capacity “to interact in an authentic communicative event utilizing different components of communicative competence extemporaneously” Brown (p. 4).

There are two types of interviews: The free interview and the controlled interview. In the free interview, no fixed set of procedures is laid down in advance. The conversation unfolds in an unstructured fashion (Weir, 2005). These interviews are in the form of extended conversations and the directions of the conversations unfold as the interview goes on. The discourse between the participants do not sound carefully formulated. The candidate has the power to change the direction of the interaction and introduce new topics. “It at least offers the possibility of the candidate managing the interaction and becoming equally involved in negotiating the meaning” (Weir, 2005, p. 154). The free conversation provides a predictable context in terms of formality, status, age of the interviewer, but the interviewer may change the context created by the attitude adopted and the role selected (Weir, 2005).

One of the disadvantages of the free interview is that “it cannot cover the range of situations candidates might find themselves in and interlocutor variables are restricted to the one interlocutor” (Weir, 2005, p. 154). On the other hand, the main advantage of this type of interview is its flexibility, which lets the interview be

modified in different aspects such as the pace, scope and level of the interaction. Weir points that a skilled interviewer can help a candidate to demonstrate their highest level of performance, with the flexibility not possible with other more structured forms of assessment.

In the controlled interview, a set of procedures is determined in advance for eliciting information (Weir, 2005). The interviewer uses an interlocutor framework of questions, instructions and prompts. He or she takes the initiative to direct the flow of the conversation and select the topic to be broached. The candidate's speech is mostly made up of his or her responses to the interviewer's questions. The interview is usually face-to-face. The interviewer usually starts to ask some personal or social questions to make the candidate feel relaxed. The controlled interview may also "enable the candidate to speak at length about familiar topics and finish at the higher levels with more evaluative routines such as speculation about future plans or the value of an intended course of study" (Weir, 2005, p. 155). One of the drawbacks of the controlled interview, Weir (1990) points out is that there is still no guarantee that the candidates will be asked the same questions in the same manner, even by the same examiner (cited in Karslı, 2002).

Weir (1990) states that one of the advantages of oral interviews is that they have a high degree of content and face validity (as cited in Karslı, 2002). Madsen (1983) also agrees that the scoring in oral interview tends to be more consistent and simple than the scoring of many guided-technique items. Therefore, they are practical for every ESL teacher to administer and score and a highly preferred means of testing speaking ability. Another advantage is that among all language examinations, oral interviews can be one of the most communicative one. In both

student-student interaction and in teacher-student interaction, candidates are highly involved in conversation. In addition, Madsen (1983) states it is remarkably flexible in terms of item types that can be included. Madsen (1983) explains the importance of this flexibility saying that the level of difficulty of items on any give interview "should vary both to maintain student confidence and the flow of the interview and also to provide an opportunity for the teacher to see how competent the student really is'' (p. 166).

On the other hand, oral interviews have some limitations. Madsen (1983) states that they are rather time consuming, particularly if taped and scored later. In addition, it is deceptively easy for it to become a simple question-and-answer session. As another drawback of oral interviews, Hughes (2003) points that "the relationship between the tester and the candidate is usually such that the candidate speaks as to a superior and is unwilling to take the initiative" (p. 119). Hence, various styles of speech cannot be elicited and many functions such as asking for information or requests for elaboration are not represented in the candidate's performance.

Weir (2005) also considers the nature of the communication in oral interviews as another drawback. Weir (2005) claims that in interviews, it is not easy to replicate features of real-life communication such as motivation, purpose and role appropriateness. He confirms his argument saying that ''the language might be purposeful but this is not always the case and there may be exchanges in which the questioner has no interest in the answer'' (p. 154). For this reason, Weir (2005) states the conversation in oral interviews have a tendency to lack real-life communication. Normally, people have conversations with each other with an interest in what the

other person is saying rather than how he is saying. Since the ways how things are said also intrude into the conversation in oral assessment, Weir (2005) states the purpose of the assessment is not in itself communicative, except of course for language teachers.

Types of Interview

Interviews are conducted either individually, in pairs, or in groups. In an individual interview, the interviewer usually takes the initiative to find out certain things about the learner and to get answers to certain questions. She has the control; the learner's speech is more or less a direct response to her questions and statements (Underhill, 1991). However; the learner still has the freedom to answer in the way he likes, or he has the chance to develop his comments and opinions. When he has finished his answer or comment, the examiner develops the topic further or introduces a new one.

With respect to the examiner's way of developing the topic or raising a new one, Luoma (2004) states that the outline the examiners have during the assessment process are very important in order to be able to structure the discourse similarly for each candidate and be fair to all examinees in this way. This outline is sometimes called an 'interlocutor frame' because it guides the interlocutor's talk in the test.

Researchers have identified a number of advantages and disadvantages to individual testing. Since the examinees are interviewed individually, it is costly in terms of examiner's time. On the other hand, as the questions that can be modified to each candidate's performance or competence, it gives the testers full control over what happens in the interview. However, this advantage may turn into a disadvantage for students.

When the interlocutor takes the initiative to control all the phases of the interaction and asks the questions, this may cause the test taker to be intimidated, which naturally prevents him from demonstrating his real oral competence. In addition, Davis (2009) identified some studies (Egyud & Glover, 2001; van Lier, 1989; Young & Milanovic, 1992) which point that power differentials between speakers and a question-and-answer style of discourse may not very well reflect actual conversation.

In addition, Lumley and Brown (1996) found that features of an interviewer's language behavior differentially support or handicap a test candidate's performance (cited in Swain, 2001). As a confirmation of this statement, McNamara and Lumley (1997) showed in their study on the effect of interlocutor and assessment mode variables that "a perception of a lack of competence on the part of the interlocutor may be interpreted as raising an issue of fairness in the mind of the rater, who may then make a sympathetic compensation to the candidate" (p. 152).

On the other hand, Luoma (2004) argues that a one-to-one test does not have to be an unstructured, interlocutor-led interview. It can also be structured and involve different tasks. Luoma (2004) shows the similarity between structured and unstructured interviews saying that "a typical structured interview would begin much like an unstructured one, with a warm-up discussion of a few easy questions, the main interaction, then, would contain the pre-planned tasks, such as describing or comparing pictures, or talking about a pre-announced or examiner-selected topic" (p. 36).

Interview in pairs have emerged as an alternative to individual interviews. The point about paired interviews is that "during the main part of the test the examinees

are asked to interact with each other, with the examiner observing rather than taking part in the interaction directly" (Luoma, 2004, p. 36). Swain (2001) mentions three arguments in favor of interviewing in pairs. First, they aim to have more types of talk than the traditional interview which will let the testers have a broader picture about the candidate's skills. Citing ffrench (2003), Davis (2009) emphasizes this advantage of paired interview saying that "paired testing may also elicit a greater variety of language functions compared to the interview format" (p. 368). The second point has to do with the relationship between testing and teaching. Swain (2001) thinks that interviews in pairs in the testing of speaking ability may promote pair works in classes or if pair work is already being practiced in the class, then the testing becomes a repetition of what is happening in teaching. Davis (2009) also states that oral communication between peers is a part of many classroom and non-classroom speaking tasks, and "so use of pair work in assessment is well suited to educational context where the pedagogical focus is fully or partially task-based" (p. 368). The third reason is economical. Since the testing is conducted in pairs, the amount of time spent decreases.

While paired interviews have many advantages, they also have many disadvantages. First, the candidates' talk may be influenced by the other person's personality, communication style and proficiency level (Luoma, 2004; Ockey, 2009). The problem is that all test takers may not have a chance to show their ability at their best. Second, testers have some concerns about the amount of responsibility that they give to the test takers (Luoma, 2004). They feel a need to make the instructions and task materials clear to the students so that they know what kind of performances will enable them to have good results.

Shohamy et al. (1986) and Fulcher (1996) state that like pair work, group interaction tasks, another commonly used format, are also generally well liked by learners (cited in Luoma, 2004). In group oral tests, there is a rater who observes and assesses each person in the group individually while they are with their fellow candidates (Van Moera, 2006). It has some advantages over the traditional type of interview. It is a resource-economical assessment because raters can assign the scores to large number of candidates in one session (Davis, 2009). This makes it useful especially for large-scale testing in institutions where a small number of teacher-raters can award scores to hundreds of students over a few hours. In addition, the examiners do not have to lead the conversation or refer to an interlocutor frame, but instead concentrate on the performance of the candidate, which makes their job easier. However, since it has some administrative concerns such as management of the size of groups and the mixture of proficiency levels in them, they are not often used in informal test of speaking (Reves, 1991, cited in Luoma (2004). On the other hand, Luoma (2004) claims that group discussions or individual presentations followed by group discussion can be quite practical.

As in pair tasks, in group interaction tasks, it is important that the task is clear to each person in the group in a way that allows each member of the entire group to participate actively. Responsibility for monitoring the progress may be challenging especially when several groups are working at the same time. Luoma (2004) suggests that the discussions might be taped, or the groups might be asked to report back to the whole class. The tapes can be used either in assessments or in self-reflection of speaking skills, by having students transcribe their section of talk, which will develop their speaking skill (Luoma, 2004). In other words, it connects assessment to

usual classroom practice. Davis (2009) states “when so much learning takes place as a result of small group collaboration and cooperation, it is arguably fairer on the students to recreate similar conditions when assessing them” (p.370).

A more recent study demonstrating the advantage of group interaction was conducted by Gan, Davison, and Hamp-Lyons (2008). In their study, they examined the production of topical talk in peer collaborative negotiation in an interactive assessment context. They applied Conversation Analysis (CA) and described and analyzed how one group of secondary ESL students focused on and constructed what they considered to be relevant to the assessment task as the interaction proceeded. In the oral discussion groups in their study, they found that the students were not only able to pursue, develop, and shift topics, but they were also able to complete the assigned task successfully. The researchers of the study concluded that such negotiation of topical talk among the students showed that peer group discussion as an oral assessment format has the potential to provide the students with the opportunities to demonstrate their real-life interactional abilities.

The Effect of Interlocutor in Oral Interviews

In each type of interview mentioned above, there is an interlocutor, rater(s) and candidate(s). Looking at each of the participants during an assessment process, we can say that the various characteristics of each have an impact on the performance of test taker. Since the primary focus of this study is the interlocutors’ behaviors during oral interview assessments, we will look closely at this affect both in paired and individual interviews. There are a number of aspects of interlocutor variability that are considered to either promote or hinder the learner’s oral competence during the assessment process.

O'Sullivan (2000) noted a critical issue related to language test tasks. "Performance in language test tasks," he said, "can be influenced by a wide range of features, which can interact unpredictably with characteristics of individual test – takers." (cited in Lumley & O'Sullivan, 2005, p. 415). One such feature which has been shown to affect learners' performance on test of spoken interaction is the gender of the person with whom they interact (O'Sullivan, 2000). In his study on exploring gender and oral proficiency interview performance, O'Sullivan (2000) looked at twelve Japanese university students (six men and six women). They were interviewed by six native speakers. All interviews were conducted in pairs. Subjects were interviewed twice, once by a woman and once by a man. On both occasions, there was an observer of the same gender as the interviewer. Video tapes of these interviews were scored and comparison of scores showed that when observer-interviewer pair was made up of women, the candidates were more likely to gain higher scores. O'Sullivan (2000) states this study supports the findings of Porter and Shen (1991) that "the gender of the interviewer is a factor which systematically affects test-candidate performance and should be controlled for in any spoken language testing situation" (p. 382)

Another factor that has been discovered to affect the learner's L2 performance in group oral discussion is the effects of group members' personalities. While pair or group tasks provide an authentic discourse for test-takers, the personal characteristics of test-takers' group members are considered as a threat to their performance by many researchers (Bonk & Ockey, 2003; Folland & Robertson, 1976).

Ockey (2009) investigated the degree to which assertive and non-assertive test takers' scores are affected by the assertiveness of their group members. The test-

takers were Japanese first-year university students. They were majoring in English as a foreign language. The Revised NEO Personality Inventory (NEO-PI-R), which assess neuroticism, extraversion, openness, conscientiousness, and agreeableness, was used. The students also took the PhonePass SET-10 (Ordinate, 2004). Two separate MANCOVA (Multivariate Analysis of Covariance) were performed on the five dependent variables: pronunciation, fluency, grammar, vocabulary, and communication strategies. One analysis was to determine to what extent the assertive test-takers' scores were affected by the levels of the assertiveness of group members. The other analysis was to find out the effect of group members' assertiveness on the non-assertive test-takers' scores.

The analyses showed that assertive test takers earned higher scores than expected when they were grouped with non-assertive test-takers. On the other hand, they got lower scores than expected when assessed with assertive test-takers. One possible explanation for this result is that the raters may have perceived the test-takers' abilities to be different based on the group in which they were assessed. That is, the raters might have compared the performance of test-takers with the performance of the group members, which made them assign different scores for the similar performances. When assertive test-takers were grouped with non-assertive ones, the raters may reward the assertive ones since they are the leader of that discussion and have more opportunities to demonstrate their oral ability. On the other hand, when all group members are assertive, it may be a disadvantage since they all compete to be leaders and have fewer opportunities to demonstrate their oral ability, which leads to the rater to penalize them. So, assertiveness may be viewed positively

when other members of the group are non-assertive, while it may be viewed negatively when all members of the group were assertive (Ockey, 2009).

The effect on pair-task performance of test-takers' acquaintanceship with their partner was viewed as an important factor in oral assessment by many researchers. O'Sullivan (2002) also conducted a study to explore this effect on students' performance. Although Porter (1991), in a study in which thirteen Arab learners were examined by known and unknown interviewers, found no evidence to support his hypothesized interlocutor-acquaintanceship effect, referring to the work of van Lier (1989) and Young & Milanovic (1992), O'Sullivan (2002) reported that "on reflection we might not expect that such an effect would be manifested (to a 'measurable' degree) in an inherently unequal interaction such as interview" (p. 279).

O'Sullivan (2002) claimed that anecdotal sources from language teachers/testers and from language learners/test-takers suggested that one's familiarity with his partner in an activity which involves interaction on a language elicitation task might positively affect performance on that task. In reviewing the effect of interlocutor acquaintanceship, O'Sullivan (2002) found that this familiarity effect has also been the focus of work in the area of psychology, where it has been suggested that "the spontaneous support offered by a friend positively affects anxiety and task performance under experimental conditions" (p.280). These studies also suggested that when a candidate is paired with a person who is considered to be a friend, they will be expected to perform better than when they are paired with a stranger.

In reviewing Katona's (1998) study of the types of meaning negotiation between Hungarian interviewers and interviewees during oral interviews, Gan,

Davison and Hamp-Lyons (2008) noted that “when the interlocutor was known to the candidate, the various negotiation sequences/exchanges represented a more natural interaction, while negotiation discourse with an unfamiliar interlocutor resulted in a more formal, artificial interaction”(p. 316).

Keeping the findings of psychology literature and anecdotes of both testers and test-takers about the relationship between acquaintanceship and better performance in mind, O’Sullivan (2002) chose a group thirty-two Japanese learners in an attempt to study this effect. The participants performed a series of three tasks (personal information exchange, narrative, and decision making). They did these tasks once with a friend, and once with a person they did not know. All performances were video-recorded and awarded scores by trained raters and transcribed for analysis.

The results of the study suggested that familiarity with one’s partner affect performance on pair-work elicitation tasks. However, the language analyses showed that there was no effect on linguistic complexity, although there was an effect on the accuracy of the language used in the different interactions. That is, O’Sullivan (2002) found that the result of his study supported the findings of the studies of Plough and Gass (1993) and Tarone and Liu (1995), which suggested that learners change their language with familiar or unfamiliar speakers.

The influence of partner proficiency is another aspect which is considered to affect the performance of test-takers in paired oral assessment. Nakatsuhura (2004) examined the discourse produced when candidates were grouped in various combinations of higher-and lower proficiency levels. She found that proficiency levels did not greatly affect conversation type, although in mixed pairs, the higher proficiency test takers spoke more.

In contrast, in another study conducted by Iwashita (1996) on the effect of proficiency level on performance, it was observed that both scores and language production differed when students were placed into equally sized groups of high and low proficiency. Iwashita (1996) tested the students once with a partner of the same proficiency and once with a partner of different proficiency (higher or lower). When students were paired with higher-proficiency partner, Iwashita (1996) found an increase in mean scores of both higher and lower proficiency students, 13% and 53% respectively.

In attempt to examine this proficiency effect, Davis (2009) conducted a study with twenty first-year students at a Chinese University. The students were divided into groups of relatively high and low English proficiency. They were tested once with a partner of similar proficiency and once with a partner of higher or lower proficiency. Discourse was analyzed using the framework "collaborative, parallel, and asymmetric" reported in Galaczi (2008). The majority of dyads produce collaborative interactions unless one examinee was paired with a much lower-level partner. In such cases, high and low levels together, the interaction tended to be asymmetric. Although the overall findings of this study (Davis, 2009), showed that the partner proficiency level had no clear effect on the measured ability levels, it was observed that lower-level examinees produced more language words when they were paired up with higher-level partners. On the other hand, higher-proficiency students showed no such effect.

In contrast to the studies above, the effect of interlocutor in a teacher-student interview is viewed quite differently by some researchers. Research concerned with teacher-student interactions demonstrates that it is the teacher who is doing most of

the talking and most of the questioning. Thus, Gan, Davison, and Hamp-Lyons (2008) claimed that the interaction is actually controlled and directed by the teacher as he or she "nominates topics, allocates turns, monitors the direction of talk, and structures the discussion" (p. 317).

In an attempt to find out these differences between paired interviews and teacher-student interviews, Brooks (2009) examined the interaction of ESL test-takers in two tests of oral proficiency: one in which they interacted with an examiner (the individual format) and one in which they interacted with another student (the paired format). She investigated how test-taker performance differed depending on whether the interlocutor was a tester or another student and also the features of interaction both in the individual and paired formats. There were eight pairs as participants and all of the test-takers participated in both test formats which included a discussion with comparable speaking prompts. The results of the study showed that looking at their test scores, the students' performance were better in paired format than when they interacted with an examiner.

Brooks (2009) also noted that "the qualitative analysis of test-takers' speaking indicated that the differences in performance in the two test formats were more marked than the scores suggested" (p. 341). When test-takers interacted with other students in the paired format, the interaction was much more complex and there was a wider range of features of interaction than in the individual format in which the testers were reliant on questions to generate interaction. She also stated that since there were more features such as prompting elaboration, referring to partner's ideas, and paraphrasing in paired interaction, the interaction was richer and more co-constructed, results which are similar to those seen by many other researchers as

fundamental aspects of oral interaction. Jacoby and Ochs observed about co-construction:

> One of the important implications for taking the postion that everything is co-constructed through interaction is that it follows that there is a distributed responsibility among interlocutors for the creation of sequential coherence, identities, meaning, and events. This means that language, discourse and their effects cannot be considered deterministically preordained by assumed constructs of individual competence (cited in McNamara, 1997, p. 456).

Another study which examines the variability of interviewer behavior and its influence on a candidate's performance was conducted by Nakatsuhara (2008). He analyzed whether there were any analytical marking categories which were especially affected by the difference in interviewer when the same candidate was interviewed by two different interviewers. In his study, two interview sessions were conducted involving two different interviewers. The speaking test lasted for twelve minutes since the primary focus of the study was interviewer-interviewee discourse. The interviews were video-taped for rating and transcribing purposes. The raters were provided with a criterion-referenced analytical scale with five marking categories: pronunciation, grammar, vocabulary resources, fluency, and interactive communication. Raters were also asked to give their reasons for awarding those scores so that "the retrospective verbal reports could help to uncover any relationship between interviewer behaviors, the candidates**'** performance, and their ratings'' (Fumiyo Nakatsuhara, 2008, p. 268).

The findings of the study showed that the two interviewers had their own ways of questioning, developing topics, reacting to the test taker's response, and these differences caused the raters to award different scores for pronunciation and fluency. That is, the study showed that there was a possible relationship between the

characteristics of interviewer behavior and particular components of language ability affected.

With respect to interlocutor variability issue in oral interviews, Lazaraton (1996) also presented a qualitative analysis of one aspect of interviewer-candidate interaction. She focused on the types of linguistic and interactional support that the native speaker interlocutor provides to the non-native speaker candidate in a one-on-one interview. In reviewing the literature, Lazaraton (1996) identified a study (Ross, 1992) which showed that interviewers routinely modified their questions during the assessment process according to the perceived difficulty the candidates were experiencing by "recompleting question turns, by suggesting alternatives to choices presented and by reformulating the questions altogether" (p. 153). Lazaraton points that the results of these studies are, on the one hand, very encouraging if the participants in oral interviews are performing conversation-like behavior and the interview approximates that real conversation as a speech event.

On the other hand, if the interviewer speech modifications are not systematic and consistent, there may add uncontrolled variability to the assessment. Lazaraton (1996) states that since the primary aim of testing outcomes is to assess the results of candidates' abilities rather than a product of context, tests should be administered using "standard, rule-governed procedures to ensure, as much as possible, that differences in scores are the result of testing environment" (p. 154). In his study on interlocutor support in oral proficiency interviews, Lazaraton (1996) collected data during one pilot CASE administration at the Suntory Cambridge English School (SCES). CASE was a two-stage oral assessment, carried out by two trained examiners with groups of six candidates. In the first stage, candidates were assessed

individually by an interlocutor and a by a second assessor who did not participate in the interaction but only observed. There was an interlocutor frame of prescribed questions which the interlocutor was to follow. In the second stage, the candidates participated in a thirteen-to fifteen-minute task-based interaction where they were assessed according to their interaction with their fellow candidates.

Results indicated that eight types of interlocutor support occurred in Phase One (interview). The types of support found were as follows: priming topics, supplying vocabulary or engaging in collaborative completions, giving evaluative responses, echoing and/or correcting responses, repeating questions with slowed speech, more pausing and definite articulation, stating question prompts as statements that merely require confirmation, drawing conclusions for candidates and rephrasing questions. Lazaraton (1996) states that these are positive findings since they suggest that processes of and practices in conversation were present in these interviews.

On the other hand, each of the examiners involved in these interviews displayed these behaviors to different extent, with some of interlocutors exhibiting these behaviors more than the others. This poses the problem of unequal opportunities for candidates to demonstrate their ability and to obtain interlocutor support. Thus, Lazaraton (1996) emphasizes that the role of examiner in the assessment process should be taken into account in the rating procedure by making the "perhaps encoded in a rating scale for Interlocutor Support" (p. 167).

The most recent study on the effect of interviewer variations on students' performance in oral interviews was conducted by Brown (2003). She specifically addressed the question of variation amongst interviewers in the ways they elicit demonstrations of communicative ability. She also examined the effect of this

variation on candidate performance and the raters' perceptions of candidate ability. In her study, she analyzed the discourse of two interviews which involve the same candidate with two different interlocutors. The study also analyzed how intimately the interviewer is involved in the construction of candidate proficiency.

The results showed that the interviewers differed in their ways of structuring sequences of topical talk, the techniques they used to question, and also the type of feedback they provided. The verbal reports of the raters explained that these differences in each interlocutor behavior led them to have different impressions of the candidate's ability. While in one interview, the candidate was considered to be more effective and eager to participate in the interaction, she was not so active and willing to show her performance in the other one. Hence, the difference in the raters' perceptions of the candidate's communicative effectiveness in the two interviews was quite marked (Brown, 2003).

Regarding the variations in the interlocutors' behaviors, Brown (2003) claims that although interviewers are usually provided with guidelines that suggest topics and general questioning focus, "specific questions are neither preformulated nor identical for each candidate; the interaction is intended to unfold in a conversational manner" (p. 1). In addition, while the unpredictability of the nature of oral interviews is favored by many researchers, who are generally proponents of oral interviews, this unpredictability is also considered as a threat for the reliability of the assessment process. For example, Stansfield and Kenyon (1992) claim that since the interviewer is free to select the topic and ask the questions he or she likes to, the same candidate may give two different performance with two different interviewers . Lazaraton (1996) warns of the dangers of such variation to fairness:

> … the achievement of consistent ratings is highly dependent on the achievement of consistent examiner conduct during the procedure, since we cannot ensure that all candidates are given the same number and kinds of opportunities to display their abilities unless oral examiners conduct themselves in similar, prescribed ways (p.166).

McNamara and Lumley (1997) emphasize the nature of oral interviews saying that:

> We must correct our view of the candidate as an isolated figure, who bears the entire brunt of the performance, this abstraction from reality conceals a potentially Kafkaesque world of others whose behavior and interpretation shape the perceived significance of the candidate's efforts but are themselves removed from focus (McNamara, 1997, p. 459).

In sum, in many studies we have seen the interlocutor effect on students' oral performance. Since interlocutors are human and fallible, some variability in interlocutor behavior will occur regardless of the level of training and standardization of behavior (Simpson, 2006). However, oral interviews which are conducted in a way which is close to real-life communication tend to reduce such an effect. The preferred nature of oral interviews, mutual interaction, is also emphasized in a study conducted by Nakamura (2008). Nakamura (2008) states that "the joy of communicating is the interaction and 'all are agents' actively participating and shaping how we talk with each other whether we realize it or not" (p. 280).

Conclusion

In the literature, it has been found that the interlocutors both in paired and teacher-student interviews have an effect on the candidates' oral performance. This effect has long been investigated but how this effect is experienced with students at the same level (Beginner) and accordingly their affective responses are not known.

In this chapter, an overview of literature on assessment, ways of speaking ability, oral interviews, and the interlocutor effect in oral interviews was presented.

The next chapter will focus on the methodology, in which the setting, participants, the instrument and the procedures of data collection and analysis will be presented.

CHAPTER III: METHODOLOGY

Introduction

The objective of this research study is to investigate interlocutor behaviors during the oral interviews through examining students' perceptions of their interlocutors' behaviors and observation of interviews. With this study, the researcher attempts to answer the following questions:

1. What are students' perceptions of their interlocutor behaviors during oral assessment interviews?

2. What are observed positive and negative behaviors of interlocutors during oral assessment interviews?

In this chapter, participants involved in the study, instruments used to collect data, data collection procedures and data analysis procedures are presented.

The Setting and Participants

The study was conducted at Karadeniz Teknik University, School of Foreign Languages, Department of Basic English. The Department of Basic English aims to provide its students with Basic English competency so that they can easily understand their courses in their departments where 30% of instruction will be in English. The department is also responsible for preparing its students to express themselves effectively in different occupational and social environments where English is used. Upon entering the school, the students are divided into three proficiency levels – beginner, pre-intermediate and intermediate – in accordance with the results of the placement test, which is given at the beginning of the semester.

The students at the preparatory school receive different amount of English instruction according to their proficiency levels. Beginner groups take 26 hours of English language courses per week, pre-intermediate groups take 24 hours, and intermediate groups take 22 hours. They study English language skills separately, so they have a final exam for each skill at the end of the year. For their speaking course, they are tested by oral interview which is administered by two instructors as interlocutors.

The participants involved in this research study are twenty English instructors who were in charge of administering oral interviews for beginning-level classes. The participants were selected for the study on the basis of willingness to participate. The twenty instructors conducted the interviews in pairs, instructors were chosen for the study only if their interviewing partner also agreed to participate. The researcher explained the process of this research study to the instructors at Karadeniz Teknik University and asked whether they would participate voluntarily in the study. Twenty of the instructors volunteered to participate in the study and signed the consent form.

In addition to this, two hundred seventy-two students from ten different classes also participated in the study. These students participated in interviews conducted by the twenty instructors who had volunteered to participate in the study. These students were asked to fill out questionnaire about the behaviors of the instructors during oral assessment interviews in which they participated. All students were in beginning-level classes. For each interview pair, two oral interviews were selected and video-recorded for the study.

Instruments

In this study, two sets of data were collected using two different types of instruments. The first set of data was collected through the questionnaire; the second data set was collected using a checklist for interlocutor behavior used during viewing of the taped interviews.

The Questionnaire

In order to find out the students' perceptions and attitudes towards their interlocutors' behaviors, a questionnaire (See Appendix A) was developed by the researcher. The questionnaire asked students to rate each of their two interviewers on separate sides of the questionnaire. Each side included two parts. In the first part, the students were asked to answer 18 items which solicited their perceptions on a five-point Likert Scale, asking them to identify behaviors of the interviewers during the oral assessment interview. Students were asked to mark each statement on the five-point scale from one ''strongly agree'' to five ''strongly disagree''. The second part of the scale had two open-ended questions. The first question asked about the situations in which the students felt bad during the oral interview. The second question asked about the reasons that made the students feel good during the oral interview.

In order to pilot the questionnaire, twenty students who were not chosen for the main study were invited to fill out the questionnaire so as to eliminate possible misunderstandings on the part of the participants. In addition to this, the questions were also checked both by ten-master-level students at Bilkent University and ten English instructors from Karadeniz Teknik University. After the answers had been

analyzed by the researcher, some of the items were revised and modified. It took the participants around 15 minutes to complete the questionnaire.

A Checklist for Interlocutor Behavior

Data were also collected in this study by the analysis of videotape recordings of twenty beginner-level students in oral interviews conducted in speaking exams in the 2009-2010 fall term with a checklist for interlocutor behavior (See Appendix C) developed by the researcher. In order to pilot the checklist, three interviews were watched by the researcher and one of her colleague. Based on this experience, the researcher made some changes on the checklist. She added sections that would allow for some parts according to the context for some specific behaviors to be noted.

During the interviews, each student was interviewed for about ten minutes, and the researcher recorded only two interviews in each class so that she could observe as many different interlocutors as possible.

After receiving permission from Preparatory School of Karadeniz Teknik University administration to collect data, the researcher talked to teachers who would administer final speaking exam, explained the study, including its aims, procedure, and future implications to the institution. Twenty out of thirty instructors were willing to participate in the study and signed the consent form. The students were also given consent forms that asked whether they would be willing to participate in the study.

Data collection procedures

Before proceeding with the research, the researcher talked to the director of Foreign Languages School at Karadeniz Teknik University to explain the purpose of the study and asked for permission for collecting data.

After receiving permission to collect data, the researcher sent a consent form to the instructors who would administer the speaking exam in the fall semester 2010. In the consent form, the purpose of the study was explained. Twenty instructors volunteered to participate in the study. The participants were not told the focus of the study in order not to be affected. The researcher only explained the process of this research study to the participants. One of the speaking teachers who was teaching beginner levels agreed to help researcher in collecting samples.

On the exam day, the researcher and her colleague went into the classes which were determined a week before video-recording. They video-recorded two interviews in each of the ten classes.

The following day, the questionnaire was administered to the classes which were video-recorded the day before. The questionnaire asked participants to rate each of the two interlocutors separately on separate parts of the questionnaire. Since 272 students rated both of their interlocutors, 544 responses gathered from the questionnaires.

After collecting the samples, the researcher designed a checklist based upon positive and negative behaviors of interlocutors gathered from the literature in order to get a more detailed analysis of interlocutor behaviors. The checklist consisted of eleven items. Using this checklist, the researcher watched three out of twenty video-recordings which were chosen randomly with one of her colleagues at Karadeniz Teknik University. Both the researcher and her colleague used the checklist for the analysis of the video-recordings by counting the number of times each interlocutor displayed that specific behavior in the checklist (See Appendix C, “A Checklist for Interlocutor Behavior”). The results from the checklist analyzed, and a high degree

of agreement was found. Both scorers had complete agreement for all behaviors for five of the six interlocutors; for one interlocutor there was a small difference in the score for one of the behaviors. Following this reliability pilot, the researcher watched and scored the rest of the recordings by herself. While watching, the researcher determined how many times the interlocutor displayed the behavior in each section of the checklist.

Data Analysis

In this survey study, the researcher used the Statistic Package for Social Sciences (SPSS) version 16.00 to analyze the data quantitatively. First, in order to analyze the frequency distribution of the participants' answers for the first 19 items of the questionnaire about specific behaviors of interlocutors, descriptive statistics were computed for each item.

The last part of the questionnaire, which included two open-ended questions, was analyzed using qualitative data analysis techniques. The data were examined through categorization of responses for each open-ended question. More specifically, the data were analyzed according to students' points of view about positive and negative behaviors of their interlocutor' behaviors. Finally, the results of the checklist were also analyzed using descriptive statistics. The data gathered from the checklist were then compared with the results of the students' questionnaires.

Conclusion

In this chapter, the setting and participants, instruments, and data collection and analysis procedures were presented. In the next chapter, the findings will be examined and the results will be discussed in detail.

CHAPTER IV: DATA ANALYSIS

Introduction

The purpose of the study was to investigate the interlocutors' observed behaviors during oral interviews. The study also aimed at to find out how students perceive their interlocutors' behaviors. Thus, the following research questions were addressed in the study:

What are students' perceptions of their interlocutors' behaviors?

What are observed positive and negative behaviors of interlocutors during oral interviews?

Two groups of subjects participated in the study. The first group of participants was 272 students, attending the Department of Basic English in the School of Foreign Languages at Karadeniz Teknik University, and the second group of participants was 20 instructors, who administered oral interviews conducted in the fall term.

In this study, two sets of data were collected through two different instruments. The first set of data was collected with a questionnaire, distributed to 272 students in total. The second set of data was collected through video-recordings of twenty interlocutors and a checklist for interlocutor behavior which was used while watching these recordings.

In this chapter, the analysis of the data is presented in two sections. In the first section, the analysis of quantitative data gathered from questionnaires will be presented. In the second section, the analysis of the qualitative data gathered from the checklist results will be presented.

Analysis of the Quantitative Data

After the completion of data collection procedures, a reliability analysis was run to calculate the Cronbach's alpha coefficient for the student questionnaire that was used in the study. Table 1 shows the result of the reliability analysis.

Table 1 - Reliability Analysis of the Questionnaire Scale used in the present study

Cronbach's Alpha	N of Items
.910	18

In order to answer the first research question, 272 beginner level students were asked to respond to a questionnaire, which had 18 questions designed on a five-point Likert scale, ranging from one ''strongly agree'' to five 'strongly disagree'' and two open-ended questions. The questionnaire was two sided. That is, the students ranked each interviewer separately, which provided 544 responses in total. Since some items were negatively worded, they were first reverse scored to calculate the reliability. Next, in order to analyze the frequency distribution of the participants' answers for each item of the questionnaire, descriptive statistics were computed for each item without reversing.

Since the items in the questionnaire have both positive and negative points about interlocutor behaviors, the frequencies will be given in two parts as negative and positive. In reporting the information out on the table, I collapsed my categories (strongly agree and agree; strongly disagree and disagree). Table 2 shows the frequencies for positive behaviors.

Table 2 - Frequencies from Student Questionnaires for Positive Interlocutor Behaviors (Individual Interlocutors)

	A/SA		Neutral		D/SD		Mean	SD
	#	%	#	%	#	%		
ITEM 1 At the beginning of the interview, the interlocutor asked me a few easy general questions.	433	83.4	26	5.0	60	11.5	1.93	.994
ITEM 2 After each question, the interlocutor gave me enough time to answer.	438	84.7	43	8.3	36	7.0	1.88	.873
ITEM 3 When I did not understand the question asked, the interlocutor did not help me by asking the same question in different ways.	110	21.3	66	12.8	341	66	1.43	1.195
ITEM 4 The interlocutor was cheerful and friendly during the interview.	370	71.7	65	12.6	81	15.7	2.12	1.141
ITEM 5 The interlocutor was not sarcastic when I could not understand or answer the question.	404	78.1	49	9.5	64	12.4	2.07	1.100
ITEM 6 The interlocutor did not try to relax me during the interview.	136	26.3	100	19.3	281	54.3	1.59	1.194
ITEM 7 The interlocutor was friendly, so I felt as if I was talking to my friend.	276	53.1	119	22.9	125	24.1	2.54	1.193
ITEM 8 When I had difficulty during the interview, the interlocutor did not help me complete my answers.	129	24.8	67	12.9	324	62.3	1.49	1.179
ITEM 9 The interlocutor did not ask the questions explicitly / in a clear way.	59	11.4	74	14.3	386	74.4	1.17	.991

ITEM 10								
Using some mimes, mimics, and body language, the interlocutor tried to explain the questions.	276	53	83	15.9	162	31.1	2.66	1.225
ITEM 11								
The interlocutor did not give extra explanations for the parts of the questions I could not understand.	90	17.4	91	17.6	337	65	1.38	1.072
ITEM 13								
The interlocutor's general manner increased my confidence in myself.	274	53.1	120	23.3	122	23.7	2.54	1.212
ITEM 14								
The interlocutor did not give positive feedback to my answers during the interview.	88	16.9	103	19.8	330	63.3	1.36	1.065
ITEM 15								
The questions the interlocutor asked were not clear and comprehensible	58	11.2	82	15.9	376	72.8	1.20	.956
ITEM 18								
The interlocutor did not generally behave in a way which showed his close interest in me and the exam.	86	16.5	60	11.5	376	72	1.21	1.167

(The five possible Likert scale answers were as follows: Strongly Agree= 1, Agree=2, Neutral=3, Disagree=4, Strongly Disagree=5)

In order to see the differences among values, they were put in the order ranging from the highest to the lowest degree. Table 3 illustrates this difference.

Table 3 - Positive Interlocutor Behaviors Ranked by Frequency from Student Questionnaires (Individual Interlocutors)

Item No.	Item	A %	DA %	Mean
2	After each question, the interlocutor gave me enough time to answer.	84.7		1.88
1	At the beginning of the interview, the interlocutor asked me a few easy general questions.	83.4		1.93

5	The interlocutor was not sarcastic when I could not understand or answer the question.	78.1		2.07
9	The interlocutor did not ask the questions explicitly / in a clear way.		74.4	1.17
15	The questions the interlocutor asked were not clear and comprehensible.		72.8	1.20
18	The interlocutor did not generally behave in a way which showed his close interest in me and the exam.		72	1.21
4	The interlocutor was cheerful and friendly during the interview.	71.7		2.12
3	When I did not understand the question asked, the interlocutor did not help me by asking the same question in different ways.		66	1.43
11	The interlocutor did not give extra explanations for the parts of the questions I could not understand.		65	1.38
14	The interlocutor did not give positive feedback to my answers during the interview.		63.3	1.36
8	When I had difficulty during the interview, the interlocutor did not help me complete my answers.		62.3	1.49
6	The interlocutor did not try to relax me during the interview.		54.3	1.59
7	The interlocutor was friendly, so I felt as if I was talking to my friend.	53.1		2.54
13	The interlocutor's general manner increased my confidence in myself.	53.1		2.51
10	Using some mimes, mimics, and body language, the interlocutor tried to explain the questions.	53		2.66

It should be noted that for several of the items (3, 6, 8, 9, 11, 14, 15, and 18), the statements were worded negatively, though still referring to positive behaviors. For these, the frequencies are reported for students who marked ‘‘Strongly Disagree’’ and ‘‘Disagree’’ to accurately compare the scores. Table 3 shows that, among the items which exemplify the positive behaviors of interlocutors, the first two items, which were giving enough time to answer after each question and asking a few general easy questions at the beginning of the interview, were agreed to by almost 85% of the participants with a mean value of between 1.88 and 1.93. As seen in the table, the third most perceived positive behavior, number five, was identified by 78.1% of the students with a mean value of 2.07. In addition, it is seen that the

following two items, which were about the way of interlocutors' asked question and the clarity of the questions, were identified at nearly the same level by students, 74.4% and 72.8% respectively. Similarly, items number 18 and 4, which were about general manner of interlocutors towards students, were identified with 72% and 71.7 % of the students. For these items, more than half of the students agreed that their interlocutors were both friendly and showed their interests in the candidates and the exam. As seen in the table, the other following five items, number 3, 11, 14, 8, and 6, which fall into the middle, neither too low or too high compared to other behaviors in the table, were identified by 66%, 65%, 63.3%, 62.3% and 54.3% of students respectively with mean values ranged between 1.43 and 1.59. Looking at these five items generally, it is seen that more than half of the students agreed that their interlocutors helped them with any kind of difficulty in understanding the questions, tried to relax them during the interview or gave explicit or gave explicit feedback to their answers. The remaining three items, number 7, 13 and 10 in the questionnaire, were identified by about 50% of the students. They were about the general manner of interlocutors and also their effort to use some mimes, mimics, and body language to explain the questions. For these items, half of the students agreed that their interlocutors were friendly, which helped them to feel at ease and have more confidence as well during the interview. The mean values for these items were calculated 2.54, 2.51 and 2.66 respectively.

The remaining 3 questions in the questionnaire, which illustrate the negative behaviors of interlocutors, are presented in two separate tables. Table 4 presents all the negative items with their frequencies, mean values, and standard deviations.

Table 4 - Frequencies from Student Questionnaires for Negative Interlocutor Behaviors (Individual Interlocutors)

	A/SA		Neutral		D/SD		Mean	SD
	#	%	#	%	#	%		
ITEM 12 The interlocutor often interrupted me while I was speaking.	37	7.2	55	10.7	423	82.1	4.03	.883
ITEM 16 The interlocutor was critical during the interview.	61	11.8	54	10.4	404	77.9	3.89	1.020
ITEM 17 During the interview, the interlocutor dealt with some other things (paper, pen, etc.) while I was speaking.	55	10.6	58	11.2	407	78.3	3.95	.969

(The five possible Likert scale answers were as follows: Strongly Agree= 1, Agree=2, Neutral=3, Disagree=4, Strongly Disagree=5)

The items which are described as negative behaviors include two specific behaviors – interrupting the student and playing with other objects during the interview- as well as a more general indicator of students' perceptions of interviewer's attitude. For all three items, the means is close to 4.0 which represents ''Disagree''. This may be interpreted as the students' perceptions of overall behavior of interviewers were generally positive. In Table 5, the frequencies are shown in the order from highest to the lowest degree.

Table 5 - Negative Interlocutor Behaviors Ranked by Frequency from Student Questionnaires (Individual Interlocutors)

Item No.	Item	A/SA %	Mean
16	The interlocutor was critical during the interview.	11.8	3.89
17	During the interview, the interlocutor dealt with some other things (paper, pen, etc.) while I was speaking.	10.6	3.95
12	The interlocutor often interrupted me while I was speaking.	7.2	4.03

Table 5 shows that the most frequent perceived negative behavior by the students is that the interviewers were critical during the interview. Of all the students, 11.8% of them agreed that their interviewers were critical during the interview. The mean value was calculated as 3.89 for this behavior. This was followed by the interviewers' dealing with some other things such as paper, pen, etc. during the interview. Ten percent of students agreed that their interlocutors displayed the incidents of this behavior. The mean value was calculated as 3.95 for this behavior. On the other hand, the negative behavior, which was identified by the smallest number of the students, 7.2%, is about interrupting. With this item, 7.2% of the students agreed that their interlocutors interrupted them while speaking, indicated by the mean value of 4.03. More than half of the students disagreed with this statement, indicated by the mean value of 3.57.

Open-ended questions in the questionnaire

Apart from these 18 Likert items on the questionnaire, there were two open-ended questions. In the first question, students were asked to write about any situation or reason which caused them to feel bad during the interview. In the second one, they were asked the reasons for feeling good during the interview. Since half of

the students (136) answered this part of the questionnaire, which meant 272 responses in total since the questionnaire was two-sided, a lot of data was not gathered from these questions. Of all the responses gathered from this part, one hundred fifty-five of them were positive while the remaining (one hundred-seventeen) were negative. However, the overall opinion from each of ten classes is as follows:

In Class one, most of the students thought that the interviewers tried to relax them both at the beginning and during the interview, which made them more productive. One of them said:

> I was really very nervous during the interview; however, I should admit that the interviewers in my exam really tried hard to make me calm. Moreover, they sometimes made jokes to make me relaxed.

They also explained that the interviewers were friendly and helpful. One student also shared the same idea saying that:

> The interviewers were very friendly and helpful, which made me feel more relaxed. They also asked the questions slowly and they tried to ask simple questions to help me when I had difficulty to speak.

In addition, they thought their interviewers listened to them attentively, which made them feel important. Another thing they liked was that the questions were asked slowly and rephrased when needed, which prevented them from misinterpreting. Moreover, the students also thought that they were given enough time to answer each question, which lessened their anxiety. One student said:

> The thing which I liked most was the interviewers were not in a hurry to ask the questions one after another, which enabled me to produce a good answer. They also gave positive feedback to my answer, which gave me more confidence in myself.

On the other hand, there were some incidents where students felt bad during the interview. They reported that one of the interviewers in a pair sometimes looked

down on them, and even had a sarcastic manner especially when they could not answer the question. The extract below presents the opinion of one student related to this issue:

> One of the interviewers was sarcastic when I did not understand the question. They sometimes laughed at what I said.

The students also were not pleased with that they were treated as if they were not beginner level but intermediate level students. One of them said:

> They behaved as if they forgot that they were speaking to beginner level students.

The most striking comment made by the students was that the interviewers caused them to lose their concentration. One of the students said that:

> When I could not answer the question asked, one of the interviewers asked in a sarcastic manner ‘‘Which question did you study?’’, which made me lose my concentration.

The students also explained that the interviewers looked angry and acted inconsiderate towards them. That is, they did not seem interested in the exam, but toyed with some other things like paper, pen, etc. Finally, most of them complained that although they started the exam with two interviewers, one of them left the room during the interview, which distracted them and caused to feel bad. The extract below illustrates the opinion of one student about this issue:

> I started the interview with two interviewers; however, later on, one of them went out in the middle of the interview, which made me feel bad since the other one was not as friendly and cheerful but seemed rather bored, nervous and less interested in what I said.

In Class 2, most of the students left this part blank, so very little data were gathered from this class. Most of the students responded as they did in their interviewers. They explained that their oral interview felt like a conversation thanks to their interviewers. One of them said:

> I liked the way my interviewers treated me. They were friendly and cheerful, which made me feel as if I was having a chat with one of my friends. Hence, I would like to have interview with the same interviewers in the spring term as well.

There were no examples for incidents which caused students to feel bad during the interview.

In Class 3, the students had similar opinions to Class 2. They thought that the interview felt like a conversation. They added that the interviewers were cheerful and friendly, which helped them feel comfortable while speaking. Another thing they liked was that the interviewers helped them when they had difficulty in understanding the questions. Moreover, the interviewers were understanding about their anxiety. Several of the students supported this idea, with one of them stating that:

> I was very nervous during the exam; however, the interviewers were very friendly and I felt they understood my anxiety and treated me accordingly.

There were some other things which were not liked by a few of the students in this class. They explained that the interviewers immediately stared to ask questions. One of them mentioned that:

> As soon as I entered the room, they just asked my name and passed on to the other questions. I wish they had asked some general questions to make me relaxed first.

Moreover, the students added that the interviewers behaved as if they did not like the answer, which made students feel that the interviewers were biased towards some ideas. One of the students supported this idea by stating that:

> I sometimes felt that I should not express my real opinions for some questions since I felt that the interviewers did not like some of my answers. This made me uncomfortable at some points during the interview.

In Class 4, the students found the interviewers friendly and understanding. They especially liked their way of asking questions. The interviewers tried to explain

the questions with examples, or they repeated when the students had difficulty. One student said:

> When the interviewers understood that I did not understand the question asked, they tried to help me giving some examples for the answer. They repeated the questions when I had difficulty to get the meaning. That is, they tried to make me feel better during the interview.

The students also stated that the interviewers did not ask questions one after the other. One of them shared this idea by stating that:

> They waited a specific time after each question, which helped me organize my ideas in my mind. They also listened to me attentively and looked interested in my answers.

On the other hand, two of the students found the duration of the interview short. They also complained that one of the interviewers was not friendly at all. One of them stated that:

> Compared to the first interviewer who started the conversation, the second one was more friendly and cheerful, which made me feel more relaxed. Hence, I think I answered better with the second one.

In Class 5, there were some positive comments but no negative ones at all since most of the students left this part blank. The students stated that their interviewers were friendly. They also tried to relax the students changing the question when they could not understand. One student wrote that:

> When I could not decide what to say or could not speak at all, the interviewers changed the topic to another topic, which helped me feel relaxed and speak well.

Interestingly, the only thing several of the students were displeased with was the fact that the interviewers sometimes had a bias towards the answers about Trabzon. One of the students said that:

> They were sensitive to my opinions about Trabzon. However, everybody does not have to love Trabzon even if it is their hometown.

Contrary to the previous class, the students in Class 6 stated no positive comments about the interviewers. Moreover, they explained that they felt pressure on them during the interview because of the teachers. One student explained that:

> During the interview, there was a kind of pressure in the room. The interviewers' tone of voice was very serious and I did not feel relaxed enough to ask for some clarification or rephrasing of the questions that I could not understand. They were not friendly.

Students also added that they did not like to be graded during the interview since they thought this behavior was distracting and caused them to lose their concentration accordingly.

Similarly, in Class 7, most of the students preferred not to answer those two items but only stated that the interviewers were friendly during the interview. One of the students said:

> Their general manner was so friendly and helpful that I felt very relaxed during the interview.

In Class 8, the students explained their satisfaction with the interviewers saying that they were quite helpful and friendly, which helped them performed better during the interview. Several of them made comments similar to this one:

> The interviewers were cheerful. They also helped me when I had any kind of difficulty during the interview.

In Class 9, the students again had all positive comments, with only one negative comment. They reported that the interviewers' attitudes were positive. They tried to help when any kind of difficulty occurred on the students' side. They were helpful both in clarifying the questions and providing them with the necessary vocabulary. With these positive attitudes, the students added that the interviewers

seemed and sounded very sincere to them. One of the student's comments was typical:

> They were so sweet and friendly. They were also very understanding even when I could not answer. In addition, when they felt that I had difficulty in understanding the question asked, both of them tried to clarify it with some examples, or stressing and repeating the key points in the question.

There were not any negative comments made by the interviewers in this class.

In Class 10, which was the last class given the questionnaire, the students had almost the same positive comments as the other classes. They explained that the interviewers were so friendly that they felt as if they were talking to a friend, which made them relaxed and feel at ease to speak. They added that their general attitude was also positive, which made them feel better. One of the students pointed that:

> The interviewers in my class were very friendly and helpful. For example, when I had difficulty in completing my sentences, they tried to help me by supplying the necessary vocabulary to complete my answer.

On the other hand, the only negative comment was that one of the interviewers was more serious than the other, which made them too nervous either to organize their sentences or think about the answers without anxiety. Another student, who shared this idea, said:

> Since one of the interviewers was quite serious compared to the other one, I felt nervous while answering her questions. Another problem was that her tone of voice was also so low that I had difficulty in hearing but was intimidated to ask for repetition because of her manner, which made her think that I did not understand the question.

Observed Behaviors of Interlocutors

In order to define the differences among interlocutor behaviors, ten interviews administered by twenty interviewers were video-recorded. To evaluate these video-recordings, a checklist for interlocutor behavior was created by the researcher based on the literature. First, three of these video-recordings were evaluated by the researcher together with one of her colleagues. Then, since the ratings of two individuals were consistent, the researcher evaluated the remaining videos by herself. After watching each video four times, means for each checklist point was calculated.

Behaviors across All Interviewers

In the checklist, there are eleven behaviors which fell into eight different categories. Table 6 below shows each checklist behavior and the criteria used to evaluate those behaviors.

Table 6 - Checklist behaviors and the criteria to evaluate the behaviors

CHECKLIST BEHAVIOR	CRITERIA
Rapport	
• General Question	• Number of general questions asked.
Supplying vocabulary or engaging in collaborative completions	
• Vocabulary	• Number of times vocabulary supplied by the interviewers when necessary.
• Completion	• Number of times helping candidate complete his answers when necessary.
Giving evaluative responses	
• Giving explicit evaluative responses	• Number of times explicit evaluative responses when given by the interviewers.
Repeating questions with slowed speech, more pausing and definite articulation	
• Speed of speech	• Number of times slowing down the speech when the candidate has difficulty.

• Stress / Pause	• Number of times stressing the words or pauses between words when the candidate has difficulty.
Stating question prompts as statements that merely require confirmation	
• Confirmation statement	• Number of times making statement which only requires confirmation instead of asking question.
Drawing conclusions for candidates	
• Stating conclusion	• Number of times stating a conclusion based on the candidate's answer
Asking questions	
• Clarification	• Number of times rephrasing the questions in a way that helps clarify it for the candidate when he / she needs.
• Giving appropriate time	• Number of times giving appropriate time to the candidate to answer each question.
Other behaviors	
• Interrupting	• Number of times interrupting the candidate when he/she speaks.

As it is seen in Table 6, there are eight main categories with subcategories. Of all eleven behaviors in the checklist, three of them, which are confirmation statement, stating conclusion and interrupting, are negative behaviors. The remaining behaviors are described as positive behaviors. In order to calculate the mean values of each behavior across the ten pairs, mean values of each item from each ten pairs were calculated. These pairs means were then added together, and then divided by ten. The resulting mean value for each behavior across the ten pairs is shown in Figure 1. While presenting the means of behaviors across all interviewers, positive behaviors will be explained first.

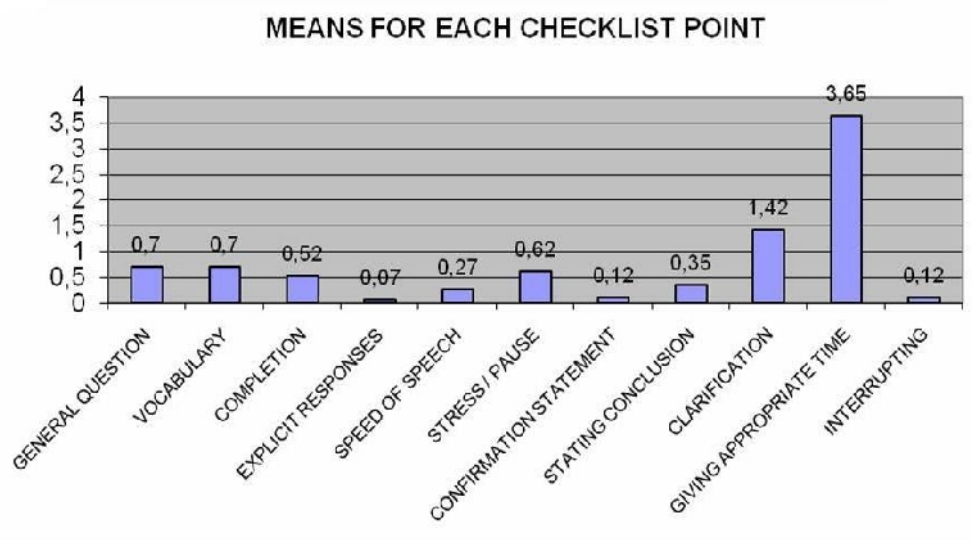


Figure 1 - Means for each checklist point

The analysis of positive behaviors in the checklist

The first category is rapport by which the interviewers were evaluated whether they ask general questions to relax the candidate. The mean for this behavior observed during the video-recordings, was 0.7 times, which means a little less than one time in every interview. The second category in the checklist is about supplying vocabulary or engaging in collaborative completions. This behavior was observed on average 0.7 times, which is the same mean value as asking general questions. On the other hand, the second behavior in this category about completing the students' answers when they have difficulty was seen about once out of two interviews on average (mean value 0.5 times).

The following section is about giving evaluative responses. It is a measurement of the number of times an instructor gives explicit evaluative responses to students' answers. This behavior was only observed in one interview for three times. The mean for this behavior was calculated as 0.75 in Pair 9 and 0.07 across the ten interviews. The next category on the checklist is about repeating questions with

slowed speech, pausing and deliberate articulation. It was seen that the interviewers slowed down their speech slightly more than once in every four interviews (0.27 times on average) while they stressed the words or paused between words when the candidate had difficulty a little more than once every two interviews (0.62 times on average).

The following category on the checklist is about the manner in which interviewers' asked questions. During the video-recordings it was observed that number of times that the interviewers rephrased the questions in a way that helped clarify them for the candidate when needed varied considerable from one to eighteen times per interview, with a mean value of 1.42 across all interviews. Similarly, they gave appropriate time to the students to answer each question at about from six to thirty –two times per interview, with a cross-interview mean of 3.65, the most frequently observed behavior.

The analysis of negative behaviors in the checklist

The first negative behavior described on the checklist is about stating question prompts as statements that merely require confirmation. The interviewers were evaluated by the extent they made statements which only require confirmation instead of asking question. In the video-recordings, it was observed that this behavior was displayed once in two pairs, twice in three pairs, and three times in two pairs while the remaining pairs did not display any incidents of this behavior (0.12).

The second negative behavior on the checklist is about drawing conclusions for candidates. It was observed that the interviewers in seven of the ten pairs displayed this behavior once, twice or three times (0.35 times on average across all pairs)

The last negative behavior is about interrupting the students while they were speaking. The mean value across ten pairs for this behavior was calculated as 0.12 times. During the video-recordings, it was observed that of all the pairs, two of them displayed this behavior two times and one pair displayed it one time across two interviews while the remaining parts did not demonstrate any incidents of this behavior.

Looking at the all of the behaviors, it can be said that the range for negative behaviors on the checklist, (making confirmation statements, stating conclusion and interrupting in the checklist), were not high compared to the positive behaviors in the checklist. The results from the observations were similar to those found from the questionnaire, specifically items number seven, confirmation statements, number eight, stating conclusion, and number eleven, interrupting.

Interviewer Behaviors for Individual Pairs

Although Figure 1 gives the overall results for each behavior in the checklist, the researcher also analyzed the behaviors for each pair separately since the interviewers worked as a team partner. This analysis allowed the comparisons of specific differences that were observed among the behaviors of different pairs.

The oral interviews were conducted by two interviewers who also rated the students. During the interviews, the interviewers had a list of questions to ask the students. After interviewing each student, both interviewers rated each student and then they averaged their ratings to give a final grade to the students.

Highest Incidence of Positive Behaviors

Although the charts for all pairs are presented in Appendix A, the following two charts present the pairs which have the highest degree of positive behaviors

displayed during the interview. The researcher calculated each behavior's result for pairs by counting the times that both interviewers displayed that specific behavior in both interview and divided by four. (as each of the behaviors for individual interviewer was rated for each of the two interviews, there were two different checklist evaluation scores for each of the interviewer).

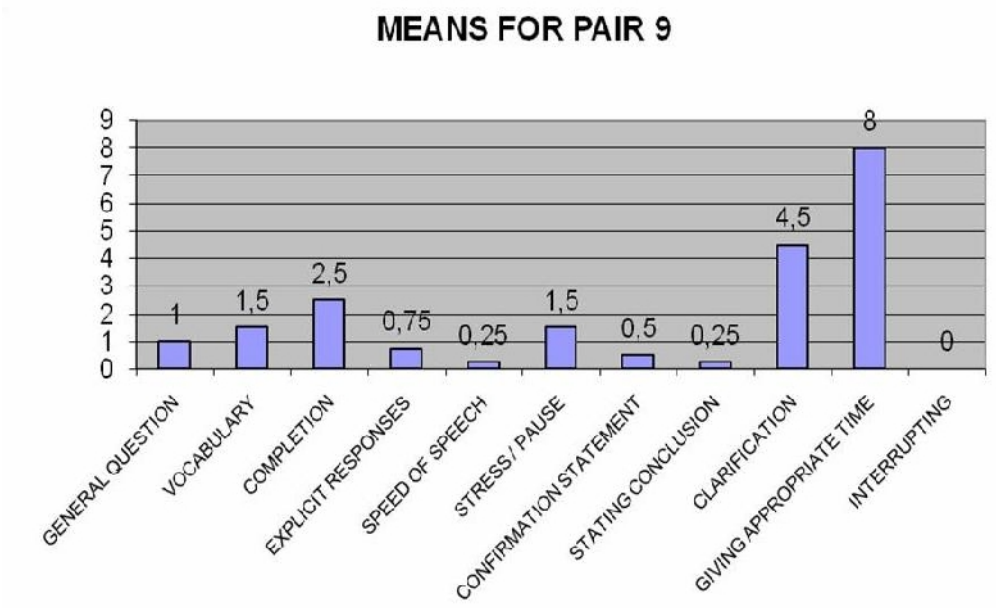


Figure 2 - Means for Pair 9

Looking at the overall results, it is seen that the means of the behaviors range from 0.25 to 8, which was not seen in any of other pair. The behavior which was observed the most frequently is number 10, giving appropriate time to the students to answer after each question. This may have resulted from the fact that this pair was the one which asked the highest number of the questions. During the video-recordings, it was also observed that the interviewers in this pair were the most helpful ones, which included asking more questions to clarify the questions for the students or waiting a sufficient amount of time to let them think about their answer or organize their ideas. On the other hand, looking at the fifth item, which is about

slowing down their speech when the candidate has difficulty, it is seen that that behavior was displayed the least. According to the notes the researcher took during the sample video-recordings, the reason for the interviewers not displaying this behavior was that the students did not need to this since the interviewers already tried to speak as clearly as possible for students to answer.

The behavior displayed at the second highest frequency, 4.5, was also about asking questions. It is seen that when the candidates needed, the interviewers rephrased the questions at a high level compared to other pairs. However, since the item measures the frequency of instructor rephrases when the candidates needed this assistance, the frequency result for this item may also reflect the frequency that the candidates for these two interviews needed this form of support.

Another behavior displayed at a relatively high level by the interviewers in this pair is about helping students complete their answers when they have difficulty during the interview. The mean for this behavior was 2.5; however, because the description of the behavior in the checklist includes the explanation ‘when needed’, the mean value for this behavior might well be higher or lower with more sample video-recordings. Behaviors Number 2, which is vocabulary and Number 6, which is stress /pause were displayed at the same rate with a frequency of 1.5. This result might also change if more than two interviewers could have been watched. However, according to the researcher’s notes during the video-recording, these interviewers displayed these behaviors whenever students needed them. As to Behavior 1 in the checklist, which is about asking general questions, the mean for this behavior is 1, which resulted from the fact that during these two interviews, since the first interviewer who started the interview already asked some general questions, there

was no need for the second interviewer to ask such questions, which was true for other interviewers in other pairs. Hence, the frequency for this behavior is usually low for all of the pairs. The other behavior which is about giving explicit responses to candidates' answers was displayed 0.75 times by the interviewers in this pair. The reason for this behavior is in fact that the interviewers were trying to show students that they were listening to them or interested in what they were saying, so they often used some expressions like ''good'', ''superb'', ''perfect'', and etc. However, this sometimes seemed to confuse the students about whether those expressions referred to the content of their answer or the correctness of the answer. The interviewers in this pair displayed no incidents of interrupting the students while they were speaking.

The interviewers in the second pair displayed incidents of positive behaviors more than any other pairs except for Pair 9.

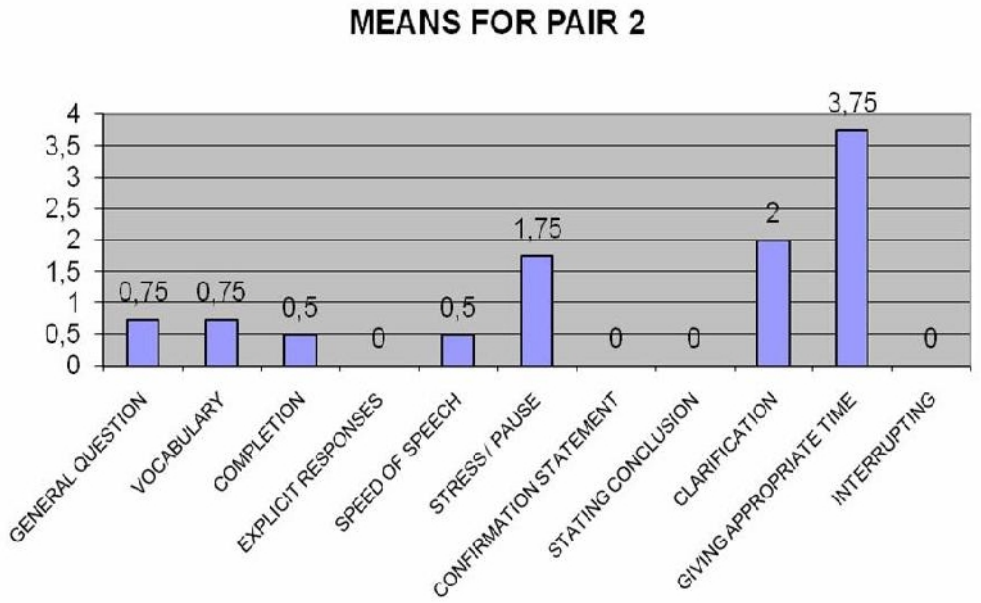


Figure 3 - Means for Pair 2

Looking at the all of the displayed behaviors in the figure above, it is seen that the interviewers in this pair gave students appropriate time to answer each

question at a high rate compared to the other behaviors, with 3.75 times on average over the two interviews. The second highest behavior, stressing the words and pausing between words when candidates need was displayed at a relatively high rate as well, which is 1.75 times. Compared to these two behaviors, some of the behaviors such as asking general questions and supplying vocabulary when necessary were not observed very often with a mean value of 0.75 times. This resulted from the fact that students did not need vocabulary supply or since one of the interviewers already asked a general question, there was no need for the other one to ask such questions, which was noted while recording the videos. Similarly, the reason why the other two behaviors, which are completing students' answers or slowing down the speech when the candidates have difficulty, were displayed 0.5 times is the fact that there were not many incidents which require the interviewers to display these behaviors.

Lowest Incidence of Positive Behaviors

Similar to the case of the pairs which showed the highest incidence of behaviors, the pairs which showed the lowest incidence of positive behaviors were examined in order to see the specific differences they showed in each behavior on the checklist. The interviewers which displayed the lowest positive behaviors were in Pair 6. The mean values of the behaviors for this pair are present in Figure 4.

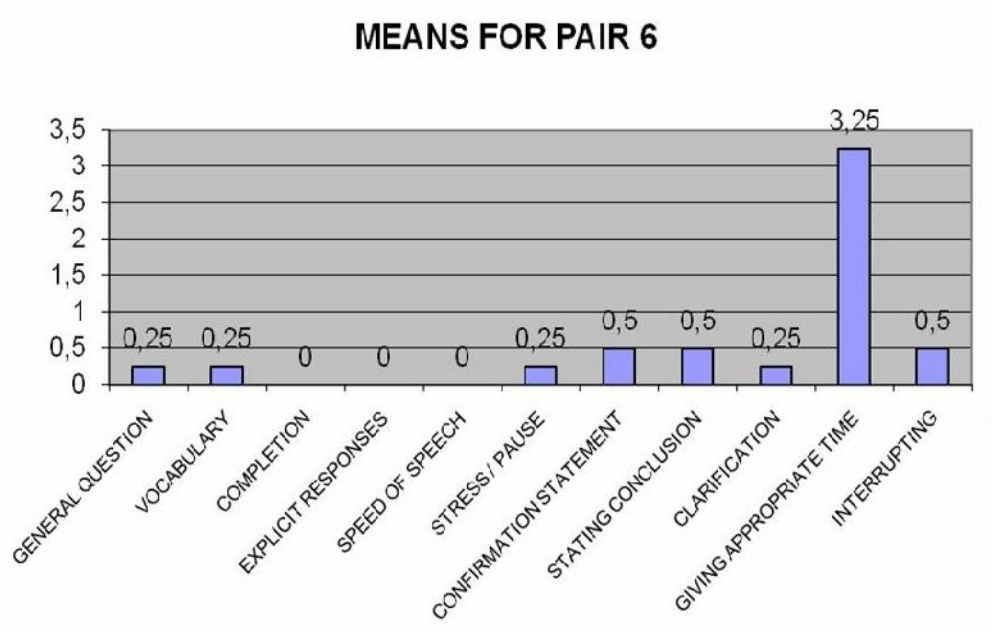


Figure 4 - Means for Pair 6

As it is seen in Figure 4, the most frequent behavior displayed by the interviewers in Pair 6 is the same as the one for other pairs, which is about giving appropriate time to students to answer each question. However; during the video-recordings it was observed that this pair did not help students by changing the questions they asked when the candidates had difficulty, but spent more time than needed on each question that the students cannot answer. The means for other positive behaviors were between 0.25 and 0.5, which is quite low compared to other pairs.

The other pair which displayed the positive behaviors on the checklist at the lowest degree is Pair 10. In Figure 5, the mean value of each positive behavior for Pair 10 will be presented.

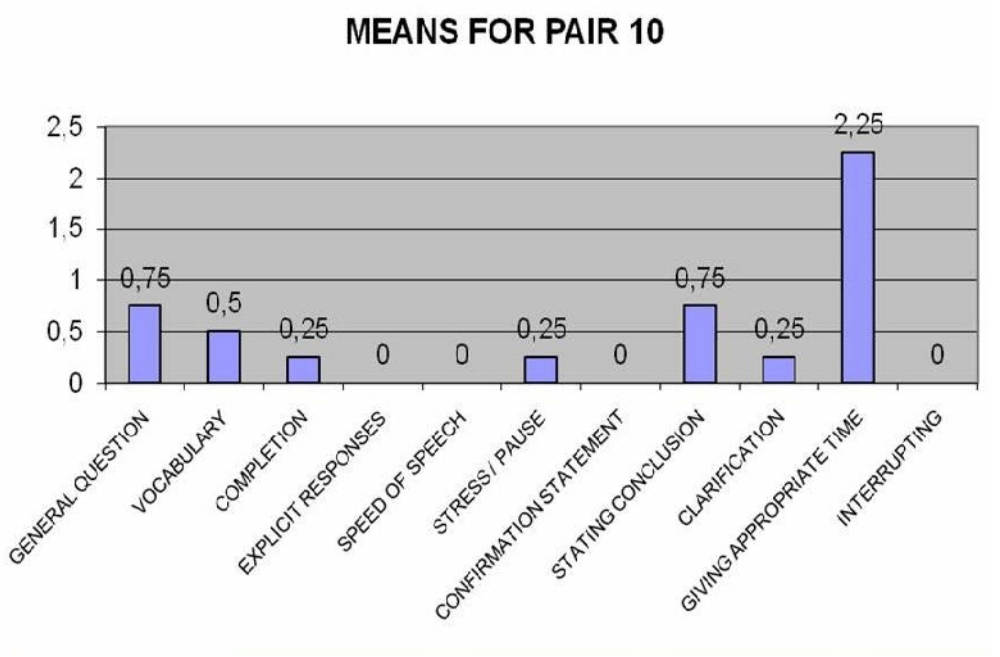


Figure 5 - Means for Pair 10

Looking at the mean values of the positive behaviors, it is seen that the interviewers displayed the tenth behavior, which is giving appropriate time to students after each question, as the highest degree, 2.25 times on average, or nine times over two interviews. Following this, it can be seen from the figure that the behaviors which were displayed at the lowest degree are completion, stress/pause, and clarification. The mean for these behaviors are 0.5, which means that the interviewers displayed these behaviors one time over two interviews. Another behavior which was displayed at a quite low level, 0.5 times on average, is about vocabulary. It was observed that the interviewers displayed this behavior two times over two interviews. The other positive behavior, which falls in the middle in the frequency order, is general question. As it is seen in Figure 5 above, the mean value for this behavior is 0.75, as the interviewers displayed this behavior three times over two interviews.

Negative Behaviors Across the Four Pairs Showing the Highest and Lowest Incidence of Positive Behaviors

Looking at the two pairs, pair 9 and pair 2, which shows the highest incidence of positive behaviors, it is seen that while the interviewers in Pair 9 displayed few incidents of negative behaviors described on the checklist, the interviewers in Pair 6 did not demonstrate any incidence of them. As it can be seen from Figure 2 above, the interviewers statements which only require confirmation 0.5 times on average, which means two times over two interviews. In addition, they also stated conclusion based on the students' answers 0.25 times on average, which means one time over two interviews.

Of the two pairs which showed the lowest incidence of positive behaviors, Pair 6 displayed the negative behaviors at a mean value of 0.5 times for each of them. That means the interviewers in Pair 6 demonstrated the incidence of each behavior two times over two interviews. As to Pair 10, the interviewers only displayed the eight behavior, which is stating conclusion based on the students' answers three times over two interviews but no incidence of other two negative behaviors.

Individual Positive Behaviors Across All Interviewer Pairs

The following section will compare ten pairs in terms of the displayed behaviors. Having organized the charts, data for each pair for each behavior were examined. The researcher analyzed positive and negative behaviors separately. The analysis of the positive behaviors will be presented first. The first positive behavior on the checklist is asking students general questions to relax them at the beginning of

the interview. Figure 6 below shows the mean values of this behavior for each pair over two interviews.

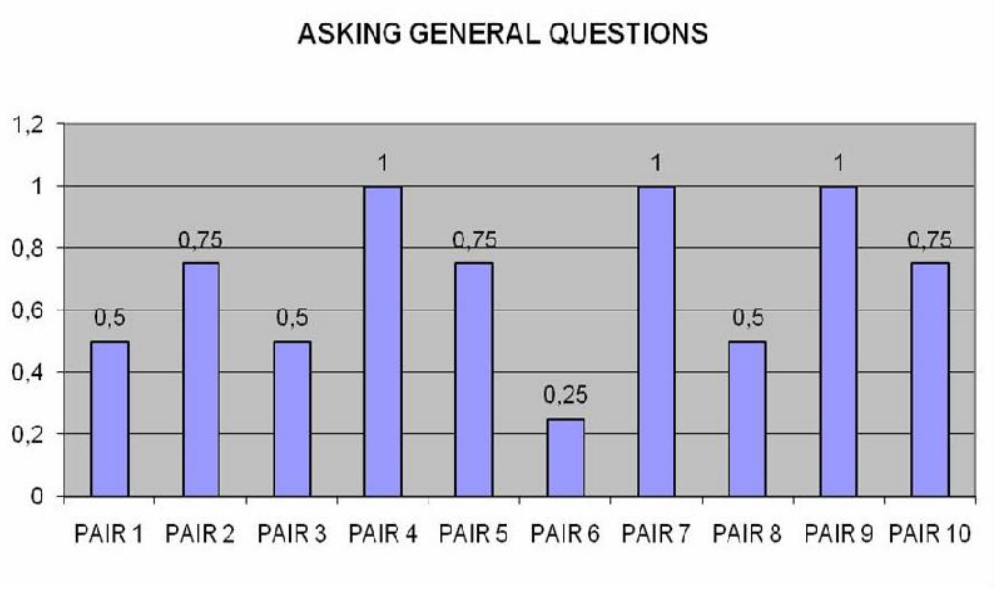


Figure 6 - General Question

Looking at the mean values of this behavior for all ten pairs, it is seen that Pair 4, Pair 7, and Pair 9 displayed this behavior at the most frequently, which is four times over two interviews. Pair 6 asked general questions one time over two interviews, which is the least among all ten interviews. The other pairs which displayed these behaviors at a similar frequency are Pair 1, Pair 2, Pair 3, Pair 5 and Pair 8, with a mean value range from 0.5 to 0.75, which means that the interviewers in these pairs asked general questions about two or three times over two interviews.

The second positive behavior on the checklist is supplying the necessary vocabulary to help the students construct an acceptable answer when necessary. Figure 7 below shows the mean values of this behavior for each ten pair over two interviews.

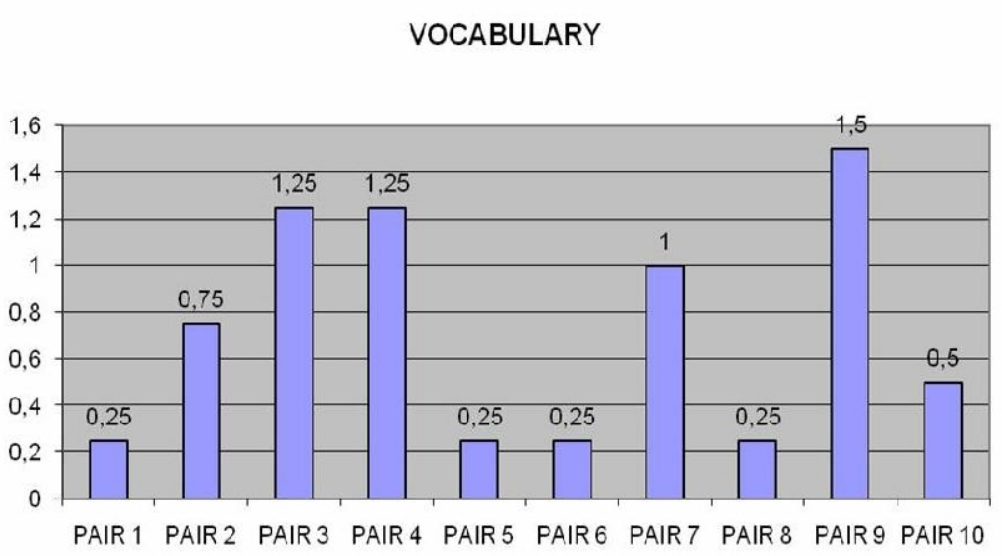


Figure 7 - Vocabulary

As it is seen in the figure above, Pair 9 is the one which helped students the most with the vocabulary when needed. It was observed during the video-recordings that the interviewers displayed this behavior six times over two interviews. This pair was followed by Pairs, 3 and 4. They helped the students with the vocabulary five times over two interviews. The Pairs, 1, 5, 6, and 8 are the ones which displayed this behavior the least, which is one time over two interviews. Pair 2 and Pair 10 displayed this behavior to a similar degree, which is three times for Pair 2 and two times for Pair 10 over two interviews.

The third behavior in the checklist measures whether the interviewers help the students complete their answers when they have difficulty during the interview. Figure 8 below shows the mean values of this behavior for individual ten pairs over two interviews.

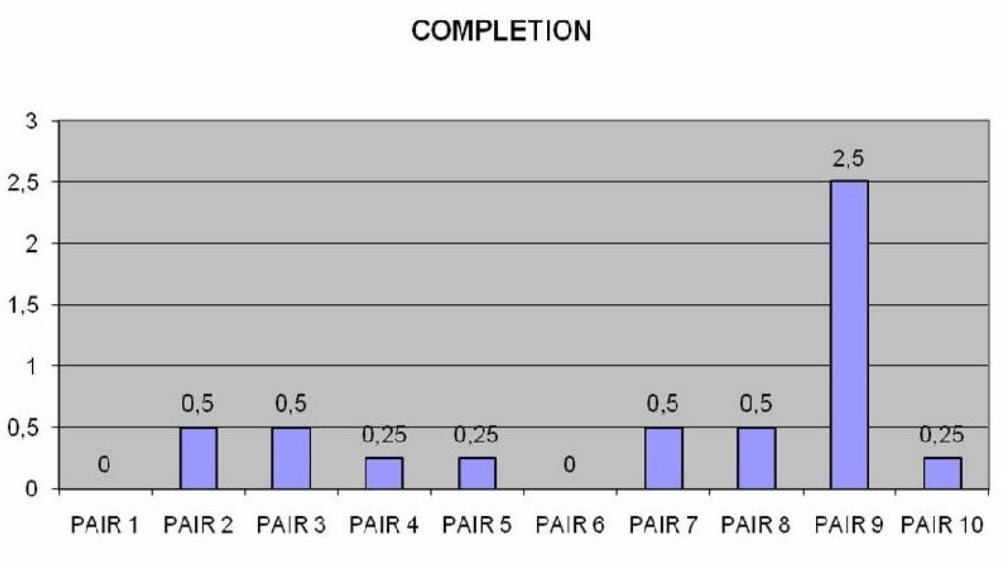


Figure 8 - Completion

Looking at this figure, it is seen that Pair 9 is the one which displayed this behavior the most. It was observed during the video-recordings that the interviewers displayed this behavior ten times over two interviews. In contrast to this pair, Pair 1 did not display any incidence of this behavior. The remaining pairs demonstrated this behavior with similar mean values change from 0.25 to 0.5, which mean one or two times over two interviews.

The fourth behavior on the checklist is giving explicit evaluative responses to students' answers. Figure 9 below shows the mean values of this behavior across ten pairs over two interviews.

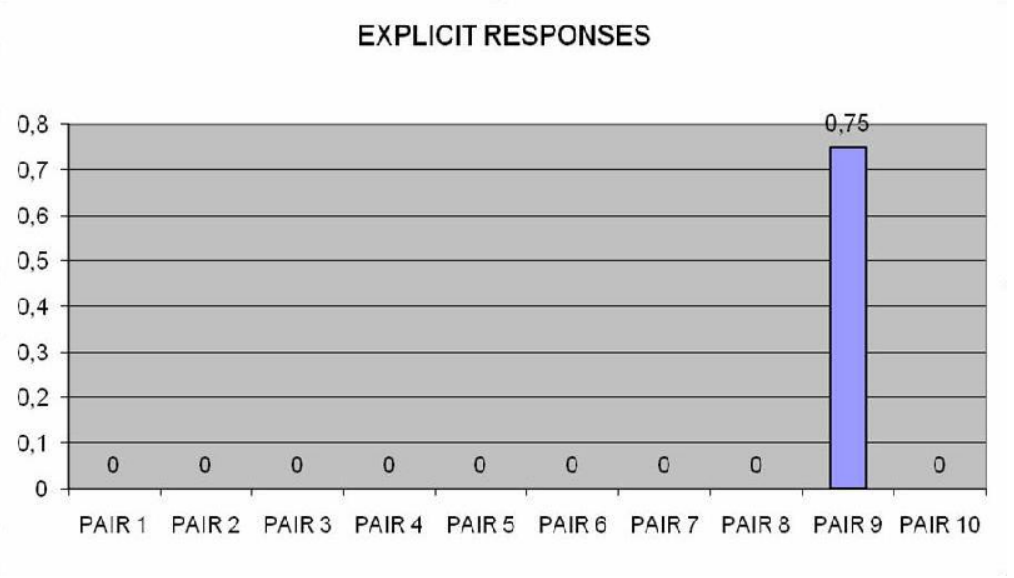


Figure 9 - Explicit Responses

As it is seen in this figure, the interviewers in Pair 9 are the only ones who gave explicit evaluative responses to students' answers, three times over two interviews.

The fifth positive behavior on the checklist measures how many times the interviewers slowed down their speech when the students had difficulty. Figure 10 shows the mean values of this behavior for each ten pairs over two interviews.

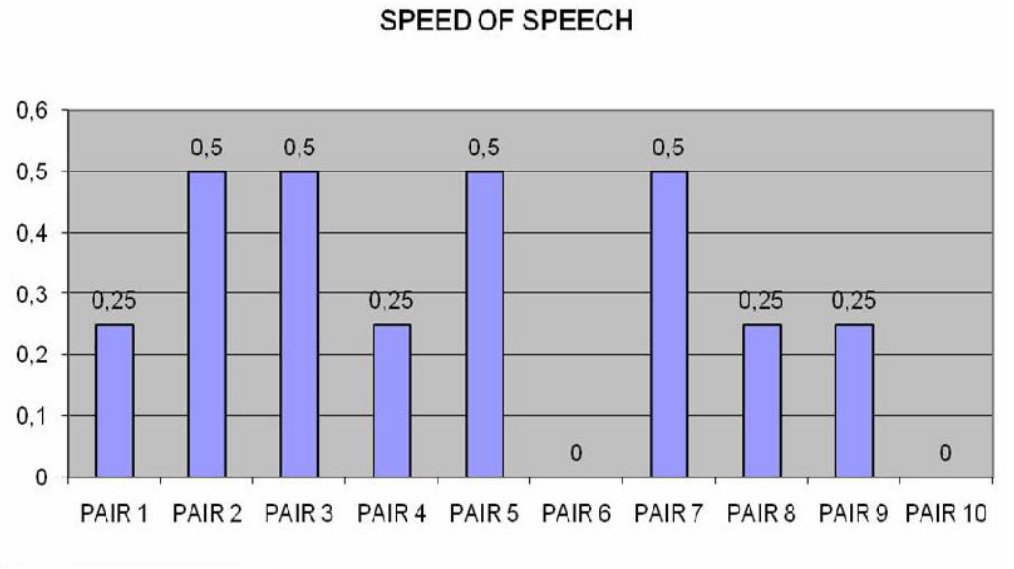


Figure 10 - Speed of Speech

Looking at the mean values of this behavior for each ten pair, it is seen that Pairs, 2, 3, 5, and 7 are the ones who displayed this behavior at a similar rate, 0.5 times on average, which means two times over two interviews. Similarly, the interviewers in the Pairs, 1, 4, 8, and 9 slowed down their speech one time over two interviews while the remaining two pairs, 6 and 10, did not display any incidence of this behavior.

The sixth positive behavior on the checklist is stressing the words or pausing between words when the students had difficulty. The mean values of this behavior for all ten pairs will be shown in Figure 11.

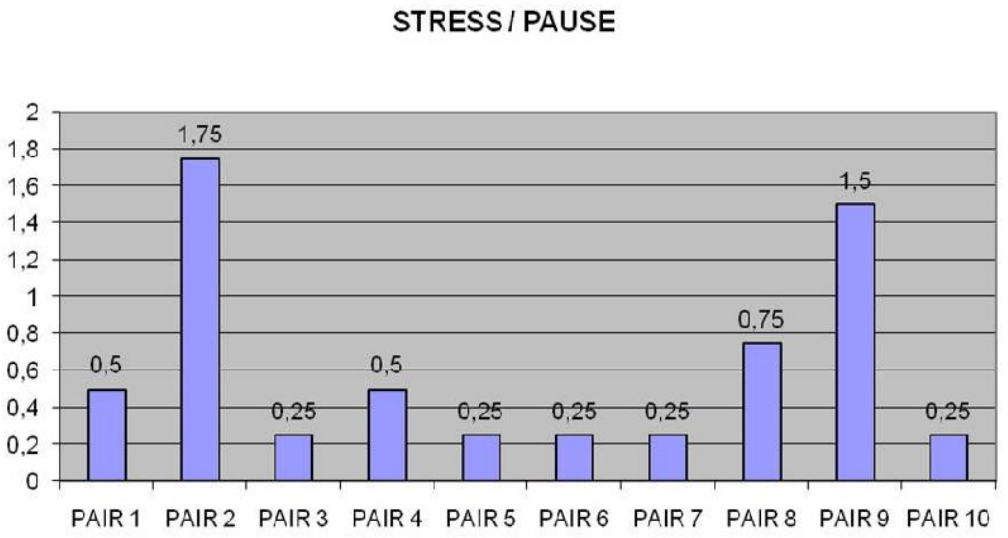


Figure 11 - Stress / Pause

As it is seen in the Figure 11, Pair 2 displayed this behavior the most with about four times on average over two interviews. Pair 9 also demonstrated this behavior to nearly the same degree as Pair 1, three times on average over two interviews. The remaining pairs displayed this behavior between 0.25 to 0.75 times, which means one, two or three times over two interviews.

The seventh positive behavior on the checklist is clarification. This behavior measures how many times the interviewers rephrased the questions in a way that helps clarify them for students when needed. The mean values of this behavior for each ten pair will be shown in Figure 12.

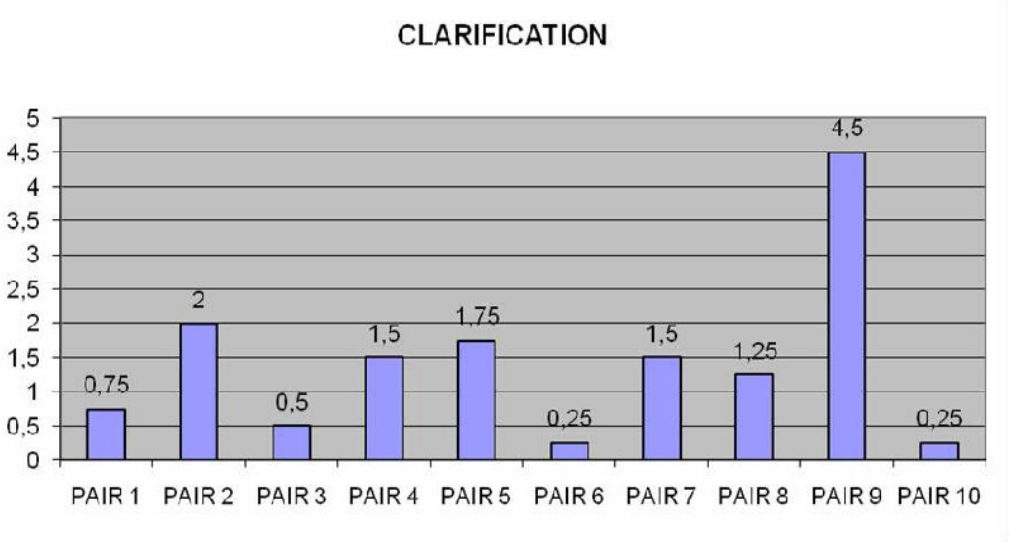


Figure 12 - .Clarification

As it is seen in the figure above, the interviewers in Pair 9 displayed this behavior eighteen times over two interviews, which is the highest degree among the other pairs. On the other hand, it was seen that Pairs 6 and 10, displayed this behavior the least with one time over two interviews. The remaining pairs demonstrated this behavior with mean values range from 0.5 to 2 times, which means between one and eight times over two interviews.

The last positive behavior in the checklist is giving appropriate time to the students after each question. The mean values of this behavior for all ten pairs will be shown in Figure 13.

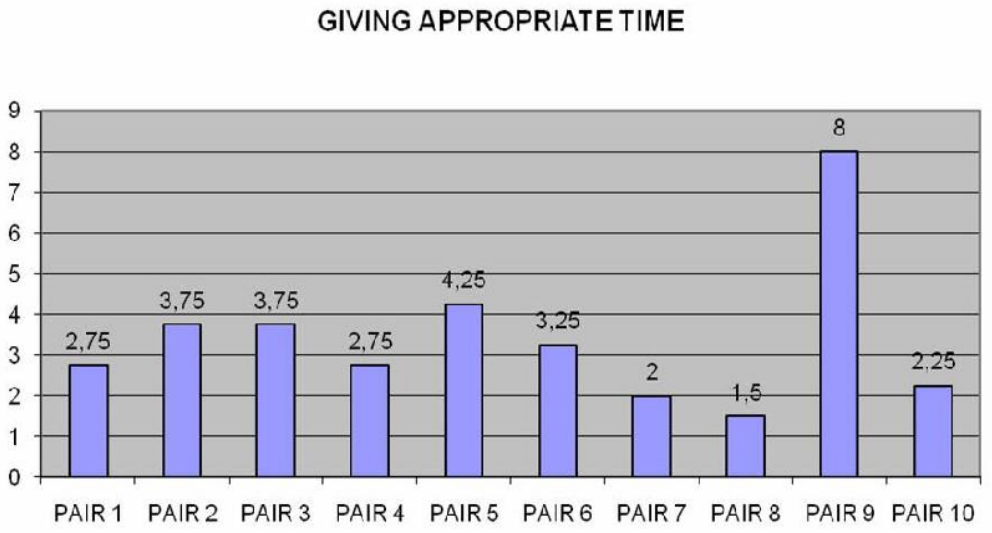


Figure 13 - Giving Appropriate Time

Looking at Figure 13, of all ten pairs, Pair 9 is the one which displayed this behavior as the most frequently with thirty-two times over two interviews. The second pair which was the closest to Pair 9 was Pair 5, which demonstrated this behavior seventeen times over two interviews. On the other hand, the interviewers in Pair 8 displayed this behavior six times over two interviews, which is the least among ten pairs. While Pairs 2, 3, and 6 showed similarity in the degree of the display of this behavior, which range from thirteen to fifteen times over two interviews, the other remaining pairs, 1, 4, 7, and 10 demonstrated this behavior between eight and eleven times over two interviews.

Individual Negative Behaviors Across All Interviewer Pairs

There are three negative behaviors described in the checklist. One of them is confirmation statement which measures whether the interviewers make statements which only require confirmation instead of asking question. The mean values of this behavior for each ten pairs will be shown in Figure 14.

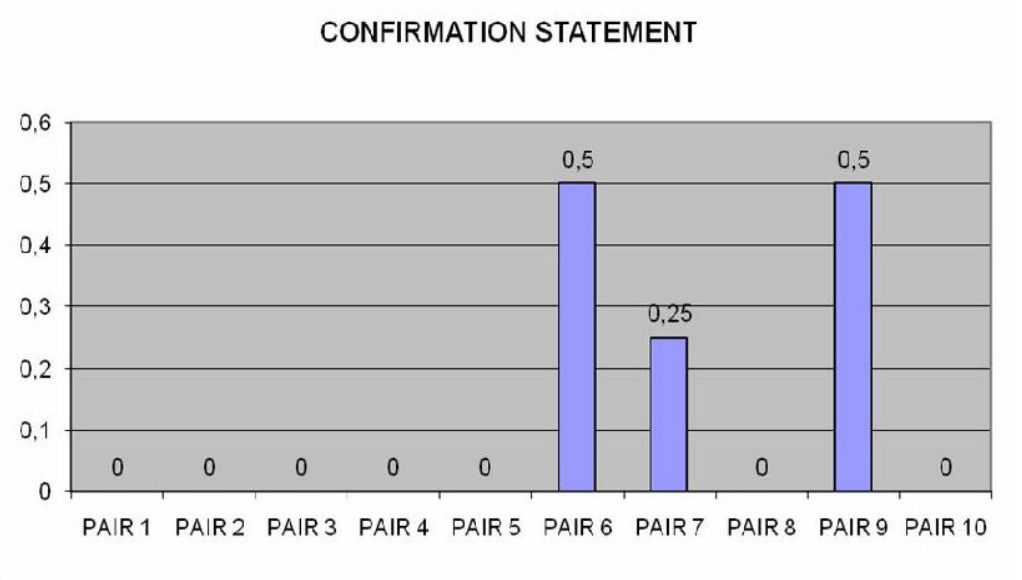


Figure 14 - Confirmation Statement

Looking at the mean values for this behavior among ten pairs , it is seen that only Pairs 6, 7, and 9 displayed this behavior with mean values range from 0.25 to 0.5, which means one or two times over two interviews.

Another negative behavior on the checklist was described as stating conclusion, which measures whether the interviewers state a conclusion based on the students' answers. The mean values for this behavior across ten pairs are shown in Figure 15.

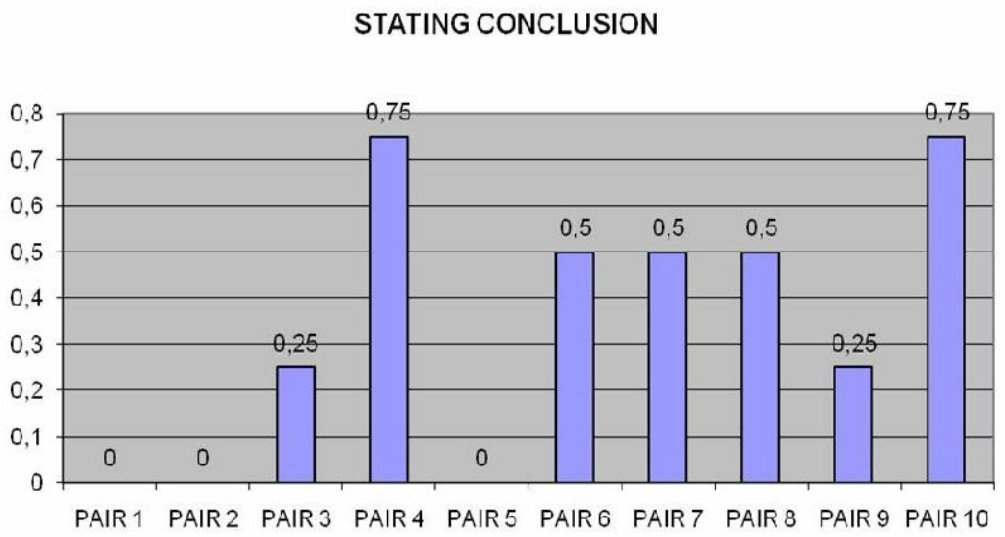


Figure 15 - Stating Conclusion

As it can be seen in the figure above, except for Pairs 1, 2, and 5, the other pairs displayed this behavior with mean values between 0.25 and 0.75. While Pairs 4 and 10 displayed this behavior three times over two interviews, which is the highest frequency among ten pairs, Pairs 6, 7, and 8 demonstrated this behavior two times over two interviews. The remaining pairs, 3 and 9 were the pairs which displayed this behavior one time over two interviews, which is the least among all ten pairs.

The last negative behavior on the checklist measures whether the interviewers interrupted the students while they were speaking. The mean values for this behavior will be shown in Figure 16.

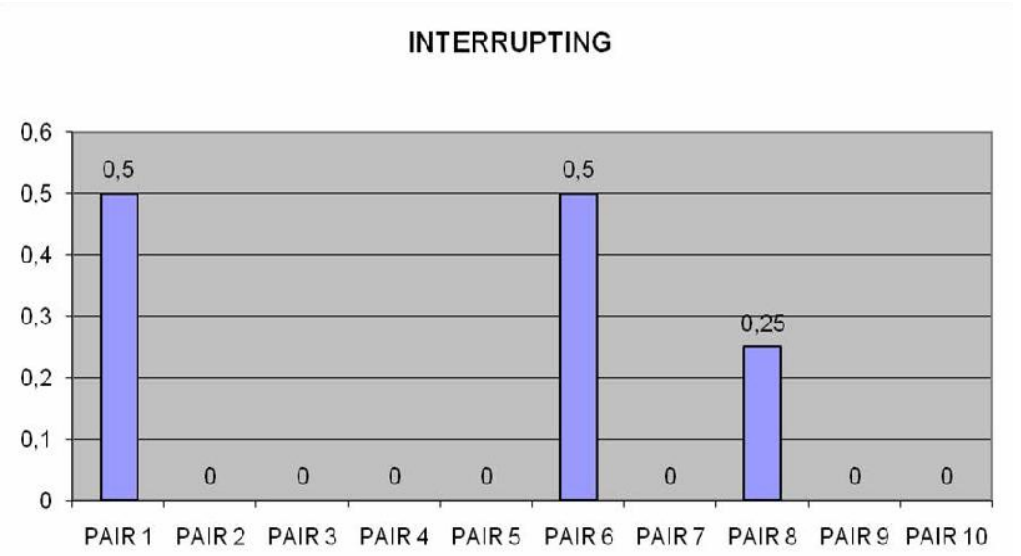


Figure 16 - Interrupting

Looking at the mean values for this behavior, it is seen that only Pairs, 1, 6, and 8 displayed this behavior, one or two times over two interviews.

Conclusion

In this chapter, the procedure of data analysis was explained and the results were reported based on the data gathered. The questionnaire results which explained the students' perceptions of their interlocutor behaviors showed that the negative behaviors displayed by the interviewers were lower compared to positive behaviors demonstrated by the interviewers. The study also analyzed the observed behaviors of interviewers across ten interviews. The results indicated that the interviewer pairs differed in the positive and negative behaviors that they displayed during the oral interviews. Moreover, it was found out that the individuals in each pair showed differences in the way they treated the students during the interview.

CHAPTER V: CONCLUSIONS

Introduction

This study aimed to investigate what were the students' perceptions of their interlocutor behaviors in their oral interviews and what positive and negative behaviors the interlocutors displayed during oral interview assessments. The data were collected through two different instruments at Karadeniz Teknik University, Department of Basic English.

First, two interviews in ten beginner level classes (for a total of twenty) which were chosen randomly were video-recorded during the oral interview assessment. The video-recordings were analyzed according to a checklist for interlocutor behavior, adapted from Lazaraton (1996). Lazaraton (1996) made a qualitative analysis of one aspect of interviewer- candidate interaction in a one-on-one interview. The analysis showed that there were at least eight types of interlocutor support in the first phase during the interviews. As the oral interviews conducted at Karadeniz Teknik University were similar, the researcher analyzed the behaviors of interlocutors observed during the interviews using this approach. Second, all the students (272) in each of those ten classes were given a questionnaire prepared by the researcher. The students identified the behaviors of their two interlocutors on the questionnaire. The questionnaire data were entered in SPSS and analyzed.

In this chapter, the research findings will be discussed and evaluated, referring to the relevant literature. The data gathered from both student questionnaire and checklist evaluation results will be used to answer the broader question ‘‘What were the interlocutors' behaviors during oral interviews?’’. In this chapter, the

limitations of the study, the pedagogical implications based on the findings, and the suggestions for further studies and overall conclusions will also be presented.

Findings and Discussion

Observed Interlocutor Behaviors and Student Perceptions

The responses of 272 participants to the questionnaire revealed that students generally perceived their interlocutors' behaviors positive. The first positive behavior which was identified by most of the students was that their interlocutor gave them enough time to answer the questions asked. Giving students appropriate time to speak or plan what they would say after a question or a given task was also emphasized by Crookes (1989). Crookes (1989), who, in examining the planning time given to the students of Japanese learners of English, found that the students produced language which was more complex and included a greater range of lexis when given more time. Similarly, Wigglesworth (1997) tried to find out whether a minimal amount of planning time, which was one minute, given to students in a test situation was adequate to have an effect on test taker's discourse. She found that candidates produced more accurate discourse especially on the more difficult tasks with this time. This behavior was also the most frequent behavior observed by the researcher during the video-recordings. Looking at the seventh category in the checklist, asking questions, it is seen that among ten pairs, the interlocutors displayed this behavior between three and eight times on average over two interviews, the highest frequency compared to other behaviors.

The second most frequent positive behavior was about the interlocutors' asking candidates a few easy general questions at the beginning of the interview. Weir (2005) also emphasized that the controlled interview, in which the interviewer

is normally standardized in the use of an interlocutor framework of questions, instructions and prompts, usually should start with personal or social questions designed to put the candidate at ease. Similarly, Madsen (1983, p. 163) mentioned the importance of establishing rapport at the beginning of oral interviews saying that ‘‘the good interviewer is neither harsh nor familiar, condescending nor intimidating. A sincere, open, supportive manner is most effective’’. He noted that everyone becomes ineffectual on any occasion when feeling threatened or talked down to. Luoma (2004) also suggested that a typical structured interview should begin with a warm-up discussion of a few easy questions such as getting to know each other or talking about the day’s events. Then, the main interaction should contain the pre-planned tasks such as describing a picture, a role-play task or a reverse interview where the examinee asks questions of the interviewer. However, during the video-recordings, it was noticed that the interview pairs did not display this behavior very often, only once or twice on average over the two interviews. This may have resulted from the fact that during the recordings, it was observed that since one of the interviewers, who started the conversation first, already asked some general questions at the beginning, there was no need for the second one to ask such questions.

The third most frequent positive behavior perceived by the students is that the interlocutors were not sarcastic towards the students when they could not understand or answer the question. The possible negative effect of using sarcasm during the interview was also emphasized by Lumley and Brown (1996). In their study, Lumley and Brown (1996) stated that the candidate is handicapped if the interlocutor uses sarcasm, interrupts, or is a passive partner in the interaction (cited in McNamara,

1997). Similarly, Brooks (2009) stated that the interlocutors may impede the test takers' performance if they interrupt the candidates or use sarcasm in the interaction.

Among the positive behaviors which was perceived the least by the students was that the interlocutors tried to explain the questions using some mimes, mimics, and body language. This positive behavior was identified by Ross and Berwick (1992). They found that examiners who used accommodations such as display questions and comprehension checks, helped the candidates in interview situations (cited in Brooks, 2009).

The other four positive behaviors, number 4, 7, 13, and 18, described in the questionnaire were about the general manner of interlocutors, which helped the candidates to feel at ease and have more confidence in themselves, were not identified by the majority of the candidates. While about 70% of the students agreed that their interlocutors were cheerful and seemed interested both in the students and the exam, about half of them (50%) stated that their interlocutors were so friendly that made them feel as if talking to one of their friends and their interlocutors general manner helped them have more confidence in themselves. Although half of the students in this study found the interlocutors to be friendly, Madsen (1983) emphasized importance of the nature of the conversation held during the oral interview. He noted that candidates should feel that they are actually talking with someone instead of simply reciting information. He added that in this way the interviewers can provide students with a genuine sense of communication.

Another positive behavior, making an effort to relax students during the interviews, was identified by a little more than half of the students. In a study conducted by Oya, Manalo and Greenwood (2004), the negative impact of anxiety on

students' oral performance was examined. They found that the anxiety had an impact on students' oral performance resulting in lower accuracy. For this reason, they suggested that instructors should facilitate a more relaxed testing environment by having an attitude which would help test takers feel at ease. In order to achieve this, they advised instructors to be lenient towards the errors students make and to take students' personality characteristics into account.

Similarly, another positive behavior identified by more than half of the students was that the interlocutors helped candidates complete their answers when they had difficulty during the interview. Swain (2001) underlined the importance of this type of assistance by stating that performance in an interview is not a solo performance but is mutual and rests on a joint construction by the participating individuals. Swain (2001, p. 280) also called the interactions in the interview a ''collaborative dialogue'' in which participants co-construct knowledge. When we look at the frequency of this support from the checklist evaluation paper, the second category, it is seen that the researcher did not observe many incidents of this behavior. In fact two pairs did not demonstrate this behavior at all. For the other eight pairs, this behavior was relatively rare, occurring on average only once in every two interviews. It should be noted that the low frequency of this behavior may be because students did not need this kind of help very much; hence, the interviewers did not need to display it so often.

Another area where interviewers seemed to demonstrate positive behaviors was in their questioning strategies. The items number 3, 9, 15, and 11 were about the way of interlocutors asked questions. A quite high number of students, 72.4%, agreed that their interlocutors asked the questions explicitly. Similarly, a quite high

number of students, 72.8%, agreed that the questions the interlocutors asked were clear and comprehensible while more than half of them, 66%, indicated that their interlocutor helped them by asking the questions in different ways. In addition, the item that points out that the interlocutors gave extra explanations for the parts of the questions students could not understand was identified by 65% of the students. Brown (2003) indicated in her study that asking unclear, closed questions and not reformulating the question when the candidate misinterpreted resulted in a poorer performance for the candidate compared to the interview conducted by another interviewer whose questions were more explicit and included more scaffolding activity. However, according to the analysis of the behavior, clarification, (the seventh category of the checklist) and the three corresponding behaviors from the questionnaire, it was observed that this scaffolding behavior was displayed as the second most frequent behavior. The ten pairs displayed this behavior on average almost one and a half times per interview. In her study, Brown (2003) examined the variation amongst two interviewers having an interview with the same candidate. She found out that the candidate performed much better with the interviewer who was more explicit in her questioning compared to other interviewer who preferred to ask closed questions or statements. Brown (2003) added that the interlocutors could make the test takers effective communicator by their scaffolding and explicit questioning and could make them poor communicators by using unclear closed questions to elicit extended responses.

Another positive behavior which was perceived by a relatively high number of students, 63.3%, is that interlocutors gave positive feedback to their answers during the interview. However, this was not observed very often by the researcher

during the video-recordings (only three times in one pair over two interviews), which is described in the third category in the checklist, giving evaluative responses. Brown (2003) also found that while one of the interviewers in her study regularly provided feedback to the test taker's responses that indicated her understanding and interest, the other interviewer tended to have rather ambiguous utterances to the test taker's responses, which appeared to confuse the test taker and create a different level of performance with each of the interviewers. The importance of giving feedback to candidates was also emphasized in the study of Nakatsuhara (2008). He mentioned that while one of the interviewers often provided comments as well as response tokens such as 'uh huh', the other interviewer gave minimal amount of feedback to the candidate's responses. He thought that the difference in types and the amount of feedback given to the test taker may have affected the perception of raters about the candidate's fluency. On the other hand, Nakatsuhara (2008) stated that the interlocutors may try to be sensitive not to interrupt the candidate's speech; hence, they may fail to fill gaps which are normally filled in during ordinary conversation, which could increase the amount of silence causing the raters to perceive that the candidate's performance is not very good.

Of all three negative behaviors, the one which was identified by the least number of candidates was interrupting. A very low percentage of students, 7.2 %, thought that they were interrupted by their interlocutors while speaking. This result is also consistent with the researcher's observation during the video-recordings. Of all ten pairs, only four pairs displayed this behavior one time on average over two interviews. Regarding interrupting, Fulcher (1996) stated that the interlocutors should be sensitive to the possibility that the students need time to organize and plan

what he/she is going to say next, and hence the amount of overlapping speech should be much less than in less formal interaction (cited in Fumiyo Nakatsuhara, 2008).

The other remaining two negative behaviors, number 16 and 17, which were identified 11% and 10% of the students respectively, emphasize the interlocutors' general manner during the interview. These items attempted to find out to what extent the interlocutors were critical during the interview or seemed interested both in the candidate and the exam rather than deal with some other things such as paper, pen, etc. According to the candidates' responses, it is seen that the interlocutors were quite sensitive in these aspects. Brown (2003) stated how the interlocutor's manner can affect the candidate's performance. In her study, the candidate performed better with the interviewer who showed keen interest in the candidate's answer by giving explicit evaluative responses or formulations that show understanding of the points made by the candidate.

The analysis of the two open ended questions also revealed positive and negative comments similar to those found in the analysis of the responses on the eighteen Likert-type items. The general positive comment of the candidates was that the interlocutors were friendly and helpful, which made them feel relaxed during the interview. A general negative comment was that the interlocutors were sometimes sarcastic towards the candidates and some of the interlocutors were tense and made the candidates more nervous during the interview. Another interesting comment was that the candidates found one of their interviewers quite helpful and friendly while they found the other rather serious and tense, which increased their anxiety and lessened oral production accordingly.

Differences among interlocutor behaviors

Looking at the differences among ten pairs for all positive and negative behaviors, it is seen that each pair sometimes showed similarities and differences in the frequency that they displayed each behavior. The first category in the checklist is rapport, asking general questions. We see that almost all pairs displayed incidents of this behavior; however, during the video-recordings details for pairs were noted by the researcher. For example, in the interview conducted by Pair 7, the interviewers were not so courteous and respectful. When the candidate sat for the interview, they said in a rude tone of voice ‘‘sign the paper and write your name’’. This contrasted with the interviewers in Pair 9, who were very understanding and tried to relax the candidates at the very beginning of the interview. For example, they did not immediately start to ask the exam questions but instead some general ones such as ‘‘Welcome. How are you? Are you excited?’’ etc. Then, they asked a preliminary question before their intended question. Lazaraton (1996, p. 158) called this behavior as ‘‘topic priming’’, which was an attempt by the interlocutor to make the intended upcoming question more clear and understandable for the candidate. She considered this kind of support very beneficial for the candidates. However, she emphasized that this sort of scaffolding can be problematic in an assessment since some interviewers display this behavior more than others, which may lead to unfairness in assessments because the candidates who have this kind of scaffolding may perform better than those who are not given this assistance. In her study, Brown (2003, p. 9) also gave some examples of this scaffolding activity, which she called ‘‘topic establishment’’. She also stated that while one of the interviewers displayed this behavior quite well, the other did not, which caused the candidate to perform worse. Harris (1969) also

underlined the importance of such kinds of questions at the beginning of the interview. He stated that the interview should begin with a few minutes of social questions such as ‘‘How are you today? What city do you come from? How long have you studied English?’’ He claimed that with this questioning, the interviewer can both put the candidates at ease and determine how well he can function in a social situation.

The second category in the checklist is supplying vocabulary or engaging in collaborative completions. The first behavior in this category is supplying the necessary vocabulary when needed. Of all ten pairs, Pair 9 is the one which helped the candidates in this sense. However, except for the situation in the interview conducted by Pair 6, for the other remaining eight pairs, the reason why the interviewers did not display this behavior is the fact that the candidates did not need this scaffolding activity. On the other hand, during the video-recordings, it was observed that the interviewers in Pair 6 did not help the candidates with the vocabulary when they had difficulty but waited and allowed a great deal of time for the candidate to find the words and complete his sentence accordingly. Harris (1969) emphasized that if it is apparent that the candidate cannot comprehend what is being said to him, the interviewer should modify his speech with some simplifications in the sentence structure and vocabulary ‘‘while making a mental note to score the candidate accordingly’’(p. 92).

The second behavior in this category is answer completion, which involves helping candidates to complete their answers when necessary. Looking at all ten pairs, it is seen that Pair 9 was again the one which helped the candidates most at this point. One possible reason that the frequency each pair displayed this behavior was

not high is that students may not have demonstrated a need for this assistance very often. In her study, Lazaraton (1996) emphasized the importance of helping students with these two points, which are vocabulary support and ‘‘collaborative completion’’ (p. 160). However, she underlined that ‘‘while these collaborative completions may actually be useful and desirable strategies in conversation, they may be counterproductive in an interview setting’’ (p. 160) if the interlocutors give fewer opportunities to candidates to talk or that interlocutor support artificially inflates the candidates’ proficiency.

The third category is giving evaluative responses to the candidates’ responses. Interestingly, Pair 9 is the only pair which gave explicit evaluative responses to the candidates’ answers. During the video-recordings, it was observed that although there was such a need in all interviews, none of the interviewers displayed this behavior except for this pair. Moreover, even in the case of Pair 9, the students actually needed more feedback from the interviewers than they received. Lazaraton (1996) pointed out that the interviewers should be very cautious in giving feedback to the candidates’ responses during the interview. She stated that if the feedback given were inconsistent and ambiguous (either to the answer or to the content of the answer), they might mislead some candidates that they were doing better in the assessment although they actually were not. Moreover, this misunderstanding may also have impact on outcome ratings.

The fourth category in the checklist is repeating questions with slowed speech, more pausing and definite articulation. There are two specific behaviors described in this category. The first one attempts to find out how often the interlocutors slowed down their speech when the candidates had difficulty. The

second behavior is stressing the words or pauses between words when the candidate has difficulty. The interviewers were observed slowing down their speech only about once every four interviews. On the other hand, they demonstrated stressing the words or pausing somewhat more. They demonstrated these behaviors a little more than once every two interviews. However, during the video-recordings, it was observed that students in some interviews might have benefited from hearing the interviewer's speech at a slower speed, especially at the very beginning of the interview. This may have resulted from the fact that since the interviewers were not the candidates' core class teachers, it was the first time they had heard that interviewer's speech, accent and way of stressing the words. Harris (1969) also emphasized this point saying that the interviewer should speak at normal conversational speed during the initial period of the interview. Regarding these two behaviors, Lazaraton (1996) stated these modifications occur after the candidates responses to interlocutor's questions. She added that if the candidates still have difficulty in understanding, the interviewer should repeat the question with ''various phonological adjustments that are characteristic of foreigner talk'' (p. 162), which would simplify the input and help him/her understand better.

The fifth category in the checklist is stating question prompts as statements that merely require confirmation, which in the context of an oral assessment, is a negative behavior. Of all ten pairs, only three of them displayed this behavior one or two times over two interviews. The problem of using statements as questions was also emphasized by Madsen (1983). He stated that it is important that the interlocutors should pause briefly and look at the candidate as if expecting an answer when they use statements as cues. Otherwise, the candidates may not feel that the

interviewer is expecting an answer. Lazaroton (1996) also agreed with Madsen on the problems of using statements as questions. She called this behavior as ‘‘presupposition’’(p. 163), which she thought may be common in daily conversation as it shows the alignment and understanding. On the other hand, she emphasized that confirmation statements may deprive candidates of opportunities to display their language ability in oral interviews, since only yes-no confirmations are required from candidates as response.

The sixth category, also another negative behavior, is drawing conclusions based on the candidates’ answer. Although seven out of ten of the pairs displayed this behavior, the frequencies showed that it was not observed very often, about one out of three interviews. Lazaraton (1996) mentioned that this behavior may be considered as an indicator of understanding on the part of the candidates; however, ‘‘ it may deprive candidates of opportunities to state for themselves what their answers imply’’(p. 164). The candidates may sometimes even have to do nothing but only agree with what their interviewers concluded, which might prevent them from speaking and distort ratings of their language. Hence, Lazaraton underlined those social practices which work well in daily conversations may not be well-suited for or may even have negative effects during oral interviews.

The seventh category concerns asking questions. There are two behaviors in this category: rephrasing the questions in a way that help clarify them for the candidates when needed and giving appropriate time to the candidates after each question. Looking at the frequency of the first behavior, clarification, it is seen that almost all pairs displayed this behavior more compared to the previous ones in the checklist, about three times every two interviews. Regarding the rephrasing the

questions to clarify them for the candidates, Lazaraton (1996) emphasized that although rephrased questions can facilitate the candidates' understanding, the interviewers should avoid multiple rephrasings, which might make the original question more difficult. She stated that the interviewers have a tendency to add increments to existing question prompts to make the question easier, even if the candidates do not ask for help, which causes trouble for them. Hence, she suggested that questions reformulations should be done carefully. The second behavior in this category is about giving appropriate time to students to answer each question. As it is seen in Figure 13, this behavior was displayed the most frequently of all behaviors on the checklist. However, it is clear that the interviewers in Pair 9 displayed this behavior at a considerably higher level compared to other nine pairs (sixteen times on average over two interviews). As Brown (2003) mentioned in her study, such differences in the opportunities given to the candidates by the interviewers to display their oral ability may lead to unfairness in the assessment. This was also noticed during the video-recordings. The interviewers in Pair 9 were trying to do more than the other interviewers to get the candidates to speak while the other pairs preferred to ask a few questions and then finish the interview.

The final category in the checklist is interrupting, another negative behavior. This measurement indicated how often the interviewers interrupted the candidates while they were speaking. It is seen that of all ten pairs, only Pair 6 and Pair 8 displayed this behavior, one or two times over two interviews. Although it seemed that both pairs displayed this behavior, the situation in each interview was quite different from each other. In the interview conducted by Pair 8, the interviewers interrupted the candidates in an attempt to help them complete their sentences, while

the interviewers in Pair 6 interrupted the students without letting them finish their sentences and ended the interview although the candidates wanted to go on their speaking. Moreover, the interviewers in Pair 6 accompanied this behavior while raising their tone of voices, which made the candidates feel more anxious and intimidated them. With respect to interrupting the candidates in the middle of the interview, Harris (1969) suggested that whatever the precise form that the conclusion takes, either satisfactory or unsatisfactory, the interviewers should take care not to give the candidate the impression that he is being cut off in the middle of a discussion and bring the interview to an end courteously. Moreover, Harris (1969) suggested that the candidate should be left with the feeling that the interview was a pleasant experience for both him and the interviewer. Lumley and Brown (1996) also discovered that the candidates are handicapped and their performance are poorer when they are interrupted by the interviewers (cited in Brooks, 2009). However, Lumley and Brown's view contradicted Nakatsuhara (2008), who claimed that if the interviewers, who wait for a long time for the candidates to go on speaking when they have difficulty, may cause the raters to perceive that the candidate is more hesitant or not very fluent in answering the questions. Hence, Nakatsuhara (2008) suggested that interrupting is necessary when applied appropriately, which means that as long as it helps to maintain the flow of the conversation.

Pedagogical Implications

In light of the findings of this study, the instructors assigned to be interlocutors during oral interviews should be aware of the fact that their behavior might differ in the way they treat the candidates. In reviewing variation in interviewer behaviors, Brown (2003) found that the research shows variation in a

variety of different behaviors. She noted that Lazaraton (1996), and McNamara (1996) found differences in the level of rapport that the interviewers establish with candidates. She also noted that Brown and Lumley (1997) and Reed and Halleck (1997) found variation in the functional and topical choices of interviewers while Lazaraton (1996) and Brown and Lumley (1997) found differences in the ways in which the interlocutors ask questions and construct prompts. Moreover, Brown (2003) also noted that Ross (1992), Ross and Berwick (1992), Lazaraton (1996) and Brown and Lumley (1997) found differences how and the extent to which interlocutors accommodate their speech to that of the candidate. Brown (2003) also noted that other researchers (Berwick and Ross, 1996) had found differences in the way in which interlocutors develop and extend topics. In this sense, the findings of this study are consistent with the study conducted by Brown (2003). That is, the interlocutors examined in this study, possessed their own way of establishing rapport, questioning, developing topics, and reacting to the candidates' responses. Considering the potential detrimental effects of these variations in the interlocutor's behaviors, some recommendations can be made.

As Weir (2005) suggested, test developers need to provide instructions in the form of an outline for the test or suggest prompts to the interlocutors or test facilitators. Weir (2005, p. 155) called this outline as ''interlocutor frame'', since it guides and constrains the interlocutor's talk and timing in the test in order to offer all candidates equal opportunity to perform to their best. She mentioned that since the interlocutor role in the test situation is to structure the discourse and to be fair to all test takers, the structure should be similar for all them. Weir (2005) also argued that the oral interview can be particularly effective as long as ''the candidates are

interviewed and assessed by both a language examiner and a subject specialist who have been standardized to agreed criteria ’’ (p. 154).

Hughes (2003) also suggested using a second tester for interviews. He mentioned that it is very difficult to conduct an interview and track of the test takers’ performance; hence, it is very helpful to have a second tester present during the interview. He believed that this person can give more attention to the test taker’s performance and have a more reliable judgment of the test taker’s performance.

Another major problem of the students during interview is that they feel very anxious and nervous because of the interlocutors’ general attitude towards them. Hughes (2003) suggested that interlocutors should show their interest in what the test takers say with verbal and non-verbal signals. He emphasized that they should be pleasant and reassuring towards the test takers. Another way of decreasing the level of test takers’ anxiety is for the interlocutors to make efforts to make the test takers not feel that they are being assessed. In order to achieve this, Hughes (2003) suggested that interlocutors should make transitions between topics and techniques as natural as possible.

As a final suggestion, Hughes (2003) emphasized that the interlocutors should be selected carefully and trained. For example, the interlocutors can be shown video-recordings of successful interviews, and then they can conduct interviews themselves which can be recorded to form the basis for critical discussion (Hughes, 2003). He pointed out that everyone does not have the natural aptitude to be a successful interlocutor. He reminded the reader that ‘‘the interlocutors need to be sympathetic and flexible characters, with a good command of the language themselves. But even

the most apt need training'' (p. 126) He emphasized that it is especially important to make the initial stages of the interview appropriate to all test takers' capacities.

As Simpson (2006) pointed out in his study,

> because interlocutors are human and fallible, some variability in interlocutor behavior will occur regardless of the level of training and standardization of behavior, hence the tension between validity and reliability is ultimately irresolvable (p. 46)

However, it would be beneficial to remember Brown's (2003) comment regarding the differences in interviewer behaviors. She pointed out that ''these differences might be taken as evidence of the natural variation that occurs amongst native speakers and therefore evidence of test validity Brown'' (p. 20)

Limitations of the Study

There are several limitations to this study that should suggest that the results be interpreted cautiously. Due to the time constraints of the study, only two interviews in each of ten classes were able to be video-recorded. Hence, the proficiency level of the candidates could not have been taken into account in the classes that participated in the study. If the researcher could have observed more interviews, the profile of the students might have showed differences and the behaviors of interlocutors may have changed accordingly.

Another limitation of the study concerns the questionnaire. Since the students were to rate their two interlocutors at the same time with one questionnaire paper, the researcher prepared it double-sided, on which the name of each interlocutor was written. However, in some classes the students did not know the names of their interlocutors, which necessitated the researcher giving explanations to assist the students. Not all students who were unsure of the identity of the interlocutors may have asked, so it is possible that some inaccurate assessments were made. Another

point about the questionnaire was that the researcher gave out the questionnaire the day after the interview. However, since students also had other final exams for other classes both on that day and throughout the whole week, the students may have been stressed while in were not so much at ease to responding to the questionnaire. This may have also been the reason for which most of students preferred to leave the open-ended part of the questionnaire blank, from which the researcher was expecting to learn much about their affective responses towards their interlocutors.

In addition to the questionnaire, additional limitations were caused by the use of checklist evaluation instrument. While analyzing the video-recordings with the checklist, the researcher noticed that some of the behaviors should have included additional details. For example, the interviewers did not display certain behaviors, the reason for each pair was quite different from each other. While in one interview, the interviewers did not demonstrate a particular positive behavior since there was not a need, for some interviews, the students would have benefited significantly from that kind of positive behavior which the interviewers did not display. To address this problem, the researcher noted details about the context for the behaviors, but these efforts may only partially address the weaknesses in the instrument. In addition, since the interviewers worked as a two-person team during the interview, the researcher could not evaluate the behaviors of individual interviewers, but had to analyze the behaviors of both interviewers by focusing upon the behaviors that either partner displayed.

A further limitation is the location of the study. The study was conducted in only one setting, the Preparatory School of Karadeniz Teknik University. While there are no obvious reasons to suggest this university differs in important ways from other

state universities in Turkey, the results about the differences in the interlocutors' behaviors and students' perceptions of these behaviors must be generalized to other settings with care.

Suggestions for Further Research

Considering the findings of the study, some suggestions can be made for further research. First, the study investigated the positive and negative behaviors of interlocutors during oral interviews and also the students' perceptions of these behaviors. Therefore, another study can be conducted to investigate to what extent these variations in the interlocutor behaviors change the students' oral performance by using both quantitative and qualitative methods. The students overall speaking competence throughout the year can be compared with the performance they displayed during the oral assessment.

Second, this study can be replicated with a larger number of interviews and interlocutors in a different EFL context in order to get a detailed picture of the variation in interlocutor behaviors.

Third, since this study was conducted in the final exam of Fall term, another study in the final exam of Spring term can be carried out to investigate if the variation in interlocutor behavior has as much impact on students' performance in the second term as the students become more proficient.

Fourth, additional studies need to examine to each pair of interlocutor behaviors, which would be a more qualitative analysis, according to the questionnaire results of their group.

Fifth, the relationship between oral interview scores and other measures of oral competency can be studied as further research.

An additional area for further research should be the development of effective observational instruments to assess interlocutor behaviors during oral interviews. As this study very clearly points out, these instruments must take into account the contextual situations that occur during oral interviews.

Conclusion

This study has revealed and supported Brown's (2003) study that the interlocutors varied in their level of rapport that they established with the candidates, the support they provided the students when they had difficulty, their reacting to the students' responses, the ways in which they asked questions, the extent to which or the ways in which they accommodated their speech to that of the candidate and the ways in which they developed and extended topics. The results of the open-ended questions in the questionnaire also showed that the students perceived their interlocutors' behaviors both positively and negatively. In the light of these findings and pedagogical implications in this chapter, interlocutors can be trained to reduce this variation in the behaviors to a minimum degree.

REFERENCES

Abbott, G., & Wingard, P. (1990). *The Teaching of English as an international language: A practical guide*. Glasgow: Collins.

Berwick, R., & Ross, S. (1996). Cross-cultural pragmatics in oral proficiency interview strategies. In M. Milanovic, Saville, N. (Ed.), *Performance testing, cognition and assessment: selected papers from the 15th Language testing research Colloquium* (pp. 34-54). Cambridge: Cambridge University Press.

Baron, J.B. & Wolf, D. P. (1996). *Performance-Based Student Assessment: Challenges and Possibilities*. Chicago: University of Chicago Press.

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

Brooks, L. (2009). Interacting in pairs in a test of oral proficiency: Co-construction a better performance. *Language Testing, 26*(3), 341-366

Brown, & Lumley, T. (1997). Interviewer variability in specific-purpose language performance tests. In V. Kohonen, Huhta, A., Kurki-Suonio, L., Luoma, S. (Ed.), *Current developments and alternatives in language assessments: proceedings of LTRC 96.* (pp. 137-150). Jyvaskyla: University of Jyvaskyla and University of Tampere.

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

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition, 11*, 367-383.

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

Egyud, G., & Glover, P. (2001). Oral testing in pairs - a secondary school perspective. *ELT Journal, 55*, 70-76.

ffrench, A. (2003). The change process at the paper level. Paper 5, Speaking. In C. Weir, Milanovic, M. (Ed.), *Continuity and innovation: Revising the Cambridge proficiency in English examination 1913-2002* (pp. 367-446). Cambridge: UCLES/Cambridge University Press.

Folland, D., & Robertson, D. (1976). Towards objectivity in group oral testing. *English Language Teaching Journal, 30*, 156-167.

Galaczi, E. D. (2008). Peer-peer interaction in a speaking test: The case of the First Certificate in English examination. *Language Assessment Quarterly, 5*, 89-119.

Gan, Z., Davison, C., & Hamp-Lyons, L. (2008). Topic Negotiation in Peer Group Oral Assessment Situations: A Conversation Analytic Approach. *Applied Linguistics, 30*(3), 315-334.

Gipps, C. (1994). *Beyond Testing*. UK: The Falmer Press.

Golstein, H., & Lewis, T. (1996). *Assessment: problems, developments, and statistical issues*. UK: John Wiley and Sons.

Harris, D. P. (1969). *Testing English as a Second Language*. United States of America: McGraw-Hill Book Company.

Hinett, K., & Thomas, J. (1999). Thinking about Self-assessment, reflection and understanding *Staffguide to self and peer assessment*. UK: The Oxford centre for staff and learning development.

Hughes, A. (2003). *Testing for Language Teachers*. Cambridge: Cambridge University Press.

Iwashita, N. (1996). The validity of the paired interview format in oral performance assessment. *Melbourne Papers in Language Testing, 5*, 1-65.

Karslı, E. S. (2002). *The inter-rater reliability of two alternative analytic grading scales for the evaluation of oral interviews at Anadolu University School of Foreign Languages.* Bilkent University, Ankara.

Katona, L. (1998). Meaning negotiation in the Hungarian Oral Proficiency Examination. In R. Young, He, A. W. (Ed.), *Talking and Testing: Discourse Approaches to the Assessment of Oral Proficiency* (pp. 240-267). Amsterdam: Benjamins.

Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: the case of CASE. *Language Testing, 13*(2), 151-167.

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

Luoma, S. (2004). *Assessing Speaking*. Cambridge: Cambridge University Press.

Madsen, H., S. (1983). *Techniques in Testing*. Oxford: Oxford University Press.

McNamara, T. F., & Lumley, T. (1997). The effect of interlocutor and assessment mode variables in overseas assessments of speaking skills in occupational settings. *Language Testing, 14*(2), 140-156.

McNamara, T.F. (1996). *Measuring second language performance*. London: Longman.

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

Nakamura, I. (2008). Understanding how teacher and student talk with each other: An exploration of how 'repair' displays the co-management of talk-in-interaction. *Language Teaching Research, 12*(2), 265-283.

Nakatsuhara, F. (2004). An investigation into conversation styles in paired speaking tests. University of Essex.

Nakatsuhara, F. (2008). Inter-interviewer variation in oral interview tests. *ELT journal, 62*(3), 266-275.

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

O'Sullivan, B. (2002). Learner acquaintanceship and oral proficiency test pair-task performance. *Language Testing, 19*(3), 277-295.

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

Oya, T., Manalo, E., & Greenwood, J. (2004). The Influence of Personality and Anxiety on the Oral Performance of Japanese Speakers of English. *Applied Cognitive Psychology, 18*, 841-855.

Plough, I., & Gass, S. M. (1993). Interlocutor and task familiarity: effects on interactional structure. In G. Crookes, Gass, S.M. (Ed.), *Tasks and language learning* (pp. 35-56). Clevedon: Multilingual Matters.

Porter, D. (1991). Affective factors in language testing. In J. C. Alderson, North, B. (Ed.), *Language testing in the 1990s. Modern English Publications in association with The British Council.* (pp. 32-40). London: Macmillan.

Reed, D. J., & Halleck, G. B. (1997). Probing above the ceiling in oral interviews: what's up there? In V. Khonen, Huhta, A., Kurki-Suonio, L., Luoma, S. (Ed.), *Current developments and alternatives in language assessments* (pp. 225-238). Jyvaskyla: University of Jyvaskyla, University of Tampere.

Ross, S. (1992). Accommodative questions in oral proficiency interviews. *Language testing, 9*, 173-186.

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

Simpson, J. (2006). Differing expectations in the assessment of the speaking skills of ESOL learners. *Linguistics and Education, 17*, 40-55.

Stansfield, C. W., & Kenyon, D. M. (1992). Research on the comparability of the oral proficiency interview and the simulated oral proficiency interview. *System, 20*, 347-364.

Swain, M. (2001). Examining dialogue: another approach to content specification and to validating inferences drawn from test scores. *Language Testing, 18*(3), 275-302.

Tarone, E., & Liu, G. Q. (1995). Situational context, variation, and second language acqusition theory. In G. Cook, Seidlhofer, B. (Ed.), *Principle and practice in applied linguistics* (pp. 107-124). Oxford: Oxford University press.

Underhill, N. (1991). *Testing Spoken Language*. Cambridge: Cambridge University Press.

van Lier, L. (1989). Reeling, writhing, drawling, strecthing and fainting in coils: Oral proficiency interviews as conversations. *TESOL Quarterly, 23*, 480-508.

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

Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Basingstoke: Palgrave Macmillan.

Wigglesworth, G. (1997). An investigation of planning time and proficiency level on oral test discourse. *Language Testing, 14*(1), 85-106.

Young, R., & Milanovic, M. (1992). Discourse variation in oral proficiency interviews. *Studies in Second Language Acquisition, 14*, 403-424.

APPENDICES

APPENDIX A: The Questionnaire

Dear Student,
I am an instructor in the Preparatory School of Karadeniz Teknik University. Currently, I am in the process of completing my Master's Degree at Bilkent University, Graduate School of Education, in Teaching English as a Foreign Language Program. I am doing a study on the perceptions of students towards their interlocutors' behaviors during oral performance assessment. Your responses will contribute to my study to a large extent. Your answers will be kept confidential. Thank you for your participation. Gülay Akın Şahin

Department: Class: Interlocutor 1 (Name and surname): Answer the following questions according to the interlocutor specified above.	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1. At the beginning of the interview, the interlocutor asked me a few easy general questions.					
2. After each question, the interlocutor gave me enough time to answer.					
3. When I did not understand the question asked, the interlocutor did not help me by asking the same question in different ways.					
4. The interlocutor was cheerful and friendly during the interview.					
5. The interlocutor was not sarcastic when I could not understand or answer the question.					
6. The interlocutor did not try to relax me during the interview.					
7. The interlocutor was friendly, so I felt as if I was talking to my friend.					
8. When I had difficulty during the interview, the interlocutor did not help me complete my answers.					
9. The interlocutor did not ask the questions explicitly / in a clear way.					
10. Using some mimes, mimics, and body language, the interlocutor tried to explain the questions.					
11. The interlocutor did not give extra explanations for the parts of the questions I could not understand.					
12. The interlocutor often interrupted me while I was speaking.					

13. The interlocutor's general manner increased my confidence in myself.					
14. The interlocutor did not give positive feedback to my answers during the interview.					
15. The questions the interlocutor asked were not clear and comprehensible.					
16. The interlocutor was critical during the interview.					
17. During the interview, the interlocutor dealt with some other things (paper, pen, etc.) while I was speaking.					
18. The interlocutor did not generally behave in a way which showed his close interest in me and the exam.					

19.……
…….
…….
……are the cases in which I felt bad during the interview.
20.……
……
……
……, hence, I felt good during the interview.

APPENDIX B: Anket

Sevgili Öğrenci,

Karadeniz Teknik Üniversitesi, Hazırlık Bölümünde öğretim elemanıyım. Şu anda Bilkent Üniversitesinde Yabancı Dil Olarak İngilizce Öğretimi (TEFL) programında yüksek lisans öğrenimi görüyorum. Karadeniz Teknik Üniversitesi Yabancı Diller Yüksekokulu Temel İngilizce Bölümünde yapılan konuşma sınavlarında, sınavı yapan öğretim elemanlarının, öğrenci performansı üzerine etkisi konusunda bir araştırma yapıyorum. Aşağıdaki sorulara vereceğiniz cevaplar yaptığım araştırmama büyük katkıda bulunacaktır. Vereceğiniz cevaplar gizli tutulacaktır.
Katkılarınız için teşekkür ediyorum.

Gülay Akın ŞAHİN

Bölüm: Sınıf: Mülakat Hocası 1 (Adı-Soyadı): Aşağıdaki soruları yukarıda yazdığınız konuşmacıyı dikkate alarak doldurunuz.	Kesinlikle katılıyorum	Katılıyorum	Kararsızım	Katılmıyorum	Kesinlikle katılmıyorum
1. Mülakatın başlangıcında hoca, genel birkaç kolay soru sordu.					
2. Hoca, her bir sorudan sonra, cevap verebilmem için bana yeterli süreyi verdi.					
3. Sorulan soruyu anlayamadığımda, hoca aynı soruyu farklı şekillerde tekrar sorarak bana yardımcı olmadı.					
4. Hoca, sınav süresince güler yüzlü bir tavır içerisindeydi.					
5. Anlayamadığım veya cevaplayamadığım sorular karşısında, hoca alaycı bir tavır sergilemedi.					
6. Hoca sınavda beni rahatlatmaya çalışmadı.					
7. Hoca sınavda dostça bir tavır sergileyerek, sanki arkadaşımla konuşuyormuşum gibi hissettirdi.					
8. Sınavda zorlandığım durumlarda, hoca cevaplarımı tamamlamamda bana yardımcı olmadı.					
9. Hoca, soruları anlaşılır bir şekilde sormadı.					
10. Hoca, sorduğu soruların açıklaması için çeşitli el, kol ve mimik hareketleriyle örneklemeye çalıştı.					
11. Sorulardaki anlayamadığım noktalarda, hoca ek açıklamalarda bulunmadı.					
12. Hoca, sınavda ben konuşurken sık sık sözümü kesti.					
13. Hocanın sınavdaki genel tavrı, kendime olan					

güvenimi artırdı.					
14. Hoca sınavda, verdiğim cevaplara pozitif tepkiler göstermedi.					
15. Hocanın sorduğu sorular net ve açık değildi.					
16. Hoca sınavda eleştiren bir tavır içerisindeydi.					
17. Mülakatta, ben konuşurken hoca başka şeylerle (kağıt, kalem, vs) ilgilendi.					
18. Hoca, genel olarak beni ve sınavı önemseyen bir tavır içerisinde değildi.					

19.……..
……
……
……durumlarında kendimi kötü hissettim.

20.……
……
……
…….olması sınav süresince kendimi iyi hissetmemi sağladı.

APPENDIX C:
A Checklist for Interlocutor Behavior

Observer: Interlocutor:

A CHECKLIST FOR INTERLOCUTOR BEHAVIOUR

Rapport

- The instructor asks general questions to relax the candidate. □□□□□□

Supplying vocabulary or engaging in collaborative completions

- The instructor supplies the necessary vocabulary to help the candidate construct an acceptable answer when necessary. □□□□□□

- When the candidate has difficulty during the interview, the instructor helps him complete his answers. □□□□□□

Giving evaluative responses

- When the instructor gives evaluative responses to students' answers, they are explicit. □□□□□□

Repeating questions with slowed speech, more pausing and definite articulation

- The instructor slows down his speech when the candidate has difficulty. □□□□□□
- The instructor stresses the words or pauses between words when the candidate has trouble. □□□□□□

Stating question prompts as statements that merely require confirmation

- The instructor makes statements which only require confirmation instead of asking question. □□□□□□

Drawing conclusions for candidates

- The instructor states a conclusion based on the candidate's answer. □□□□□□

Asking questions

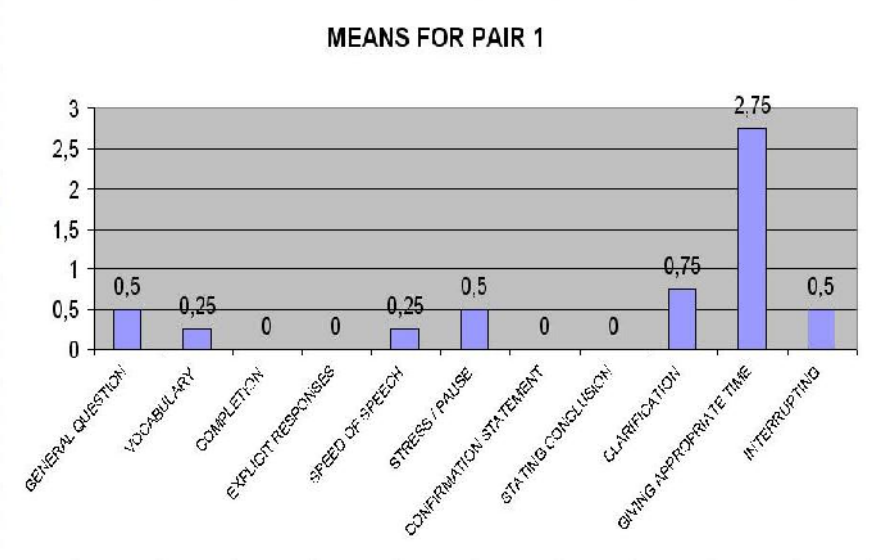
- The instructor rephrases the questions in a way that helps clarify it for the candidate when he/she needs. □□□□□□

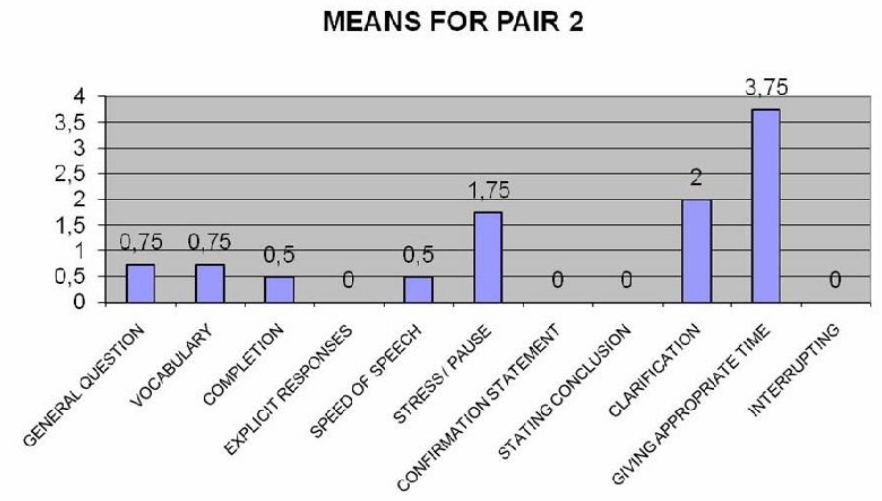
- The instructor gives appropriate time to the student to answer each question. □□□□□□

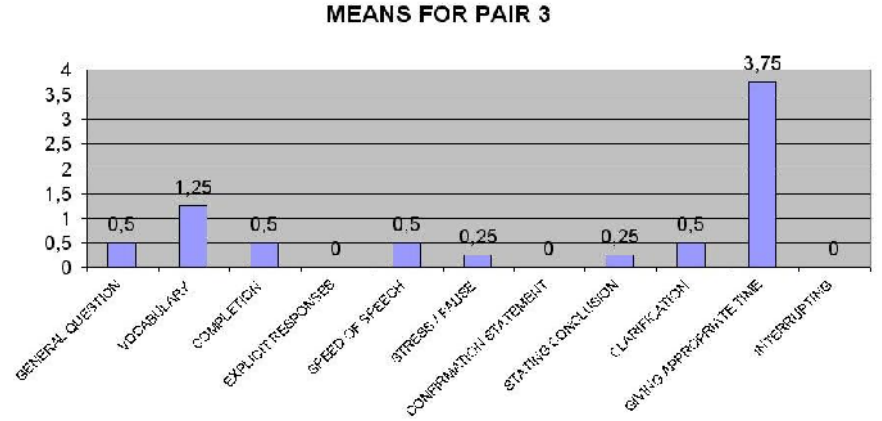
Other Behaviors

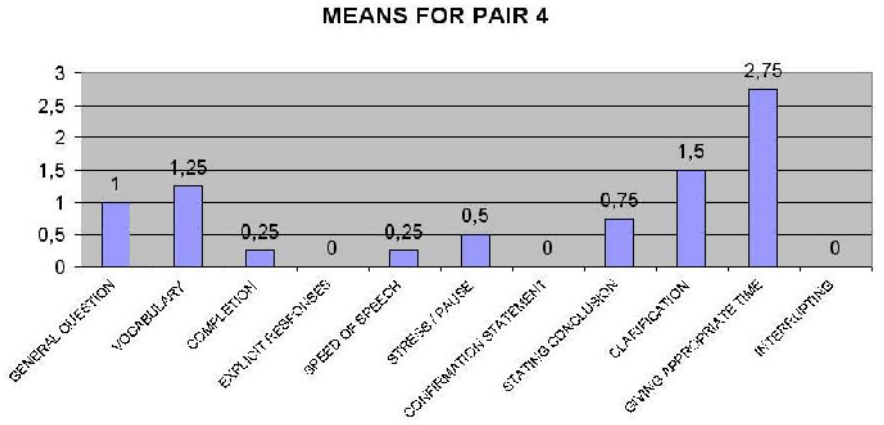
- The instructor interrupts the student while speaking. □□□□□□

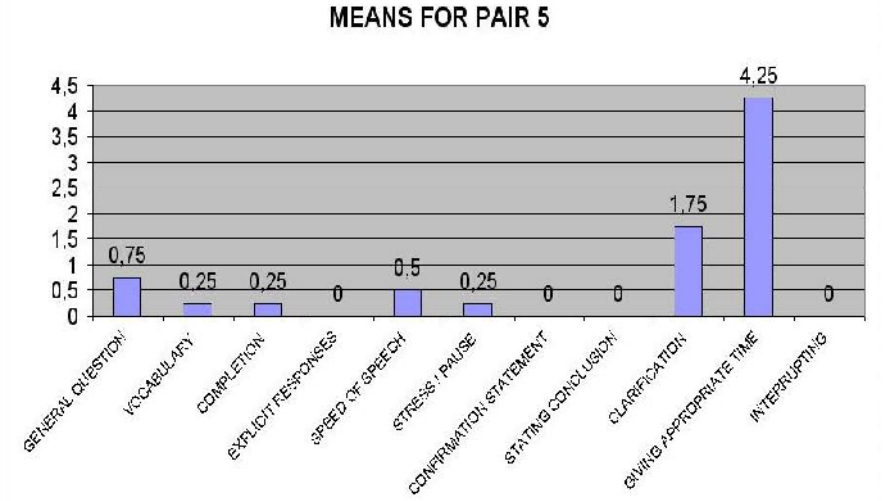
APPENDIX D: Figures of Means for Interlocutor Pair

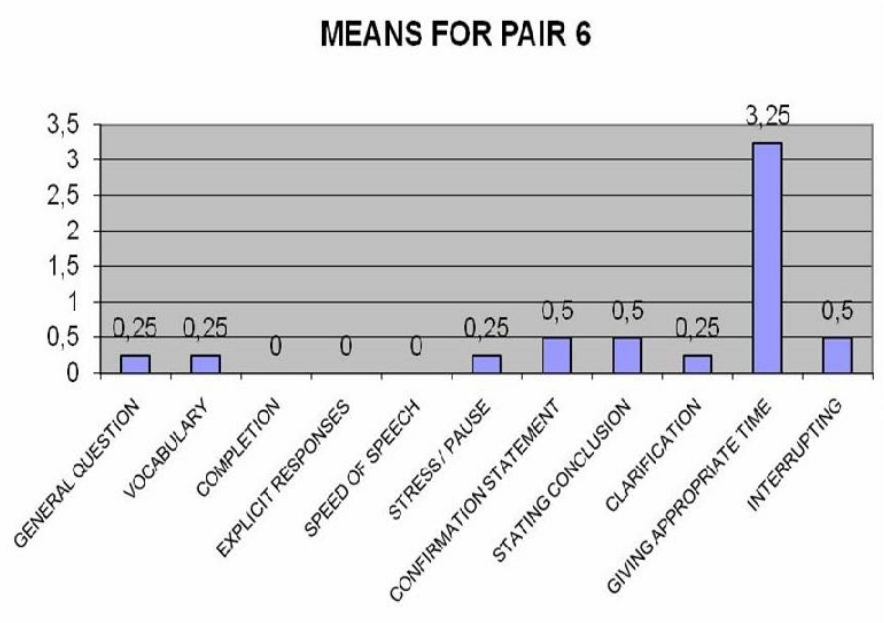




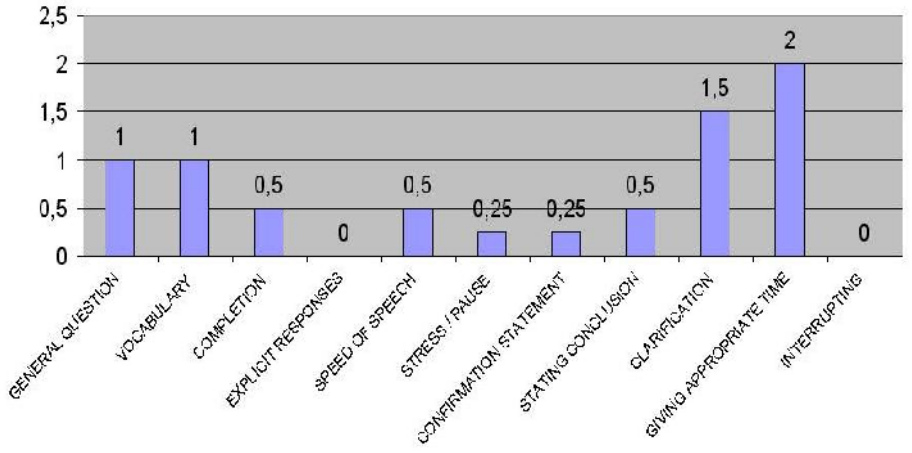



MEANS FOR PAIR 4
3
2,5
2
1,5
1
0,5
0
1
1,25
0,25
0
0,25
0,5
0
0,75
1,5
2,75
0
GENERAL QUESTION
VOCABULARY
COMPLETION
EXPLICIT RESPONSES
SPEED OF SPEECH
STRESS / PAUSE
CONFIRMATION STATEMENT
STATING CONCLUSION
CLARIFICATION
GIVING APPROPRIATE TIME
INTERRUPTING

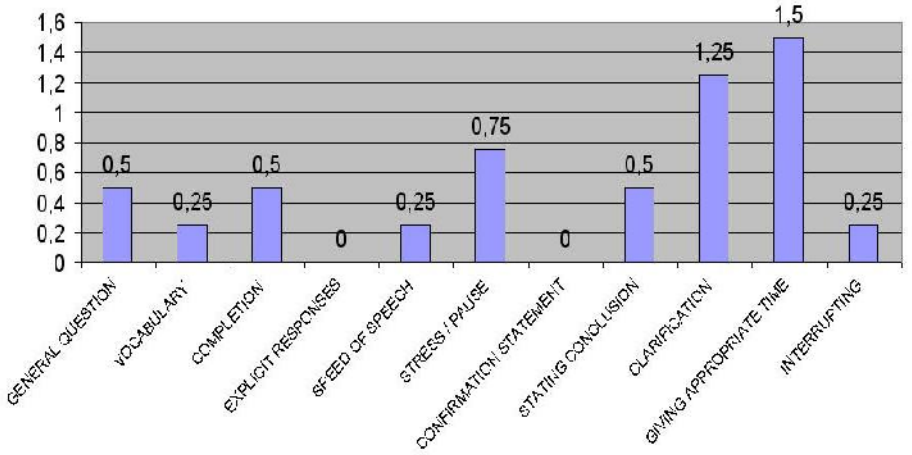

MEANS FOR PAIR 5
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
0,75
0,25
0,25
0
0,5
0,25
0
0
1,75
4,25
0
GENERAL QUESTION
VOCABULARY
COMPLETION
EXPLICIT RESPONSES
SPEED OF SPEECH
STRESS / PAUSE
CONFIRMATION STATEMENT
STATING CONCLUSION
CLARIFICATION
GIVING APPROPRIATE TIME
INTERRUPTING

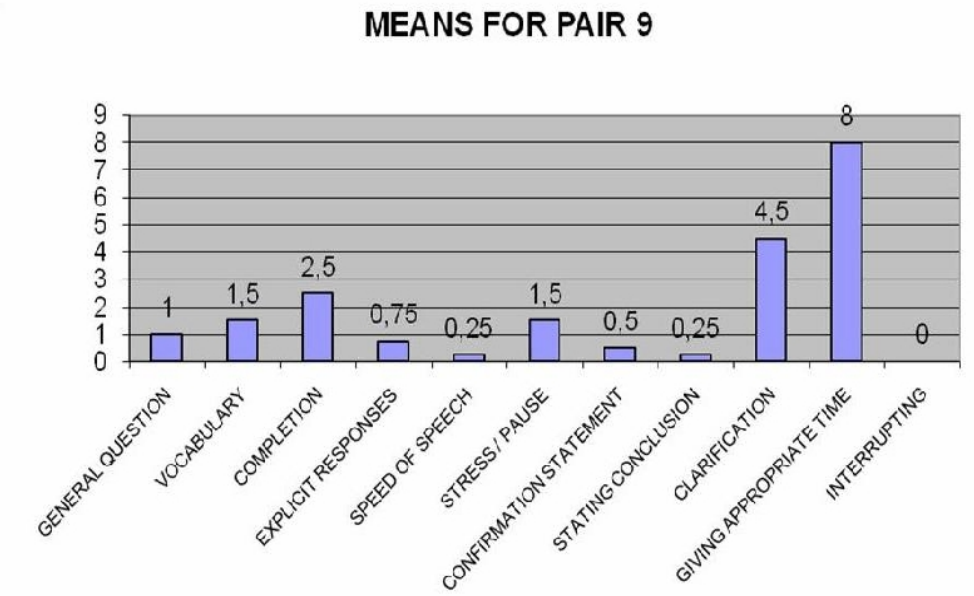

MEANS FOR PAIR 6
3,5
3
2,5
2
1,5
1
0,5
0
0,25
0,25
0
0
0
0,25
0,5
0,5
0,25
3,25
0,5
GENERAL QUESTION
VOCABULARY
COMPLETION
EXPLICIT RESPONSES
SPEED OF SPEECH
STRESS / PAUSE
CONFIRMATION STATEMENT
STATING CONCLUSION
CLARIFICATION
GIVING APPROPRIATE TIME
INTERRUPTING

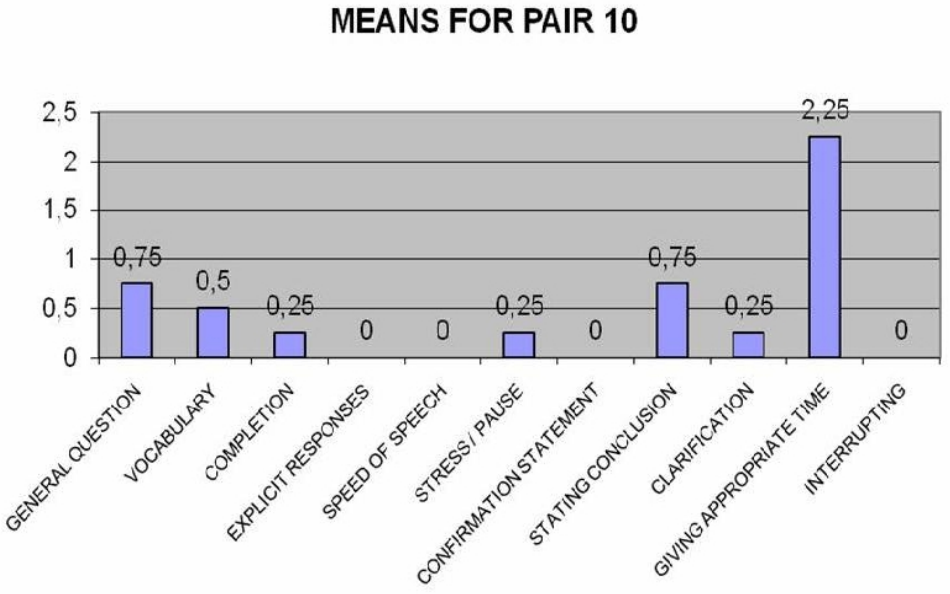
MEANS FOR PAIR 7



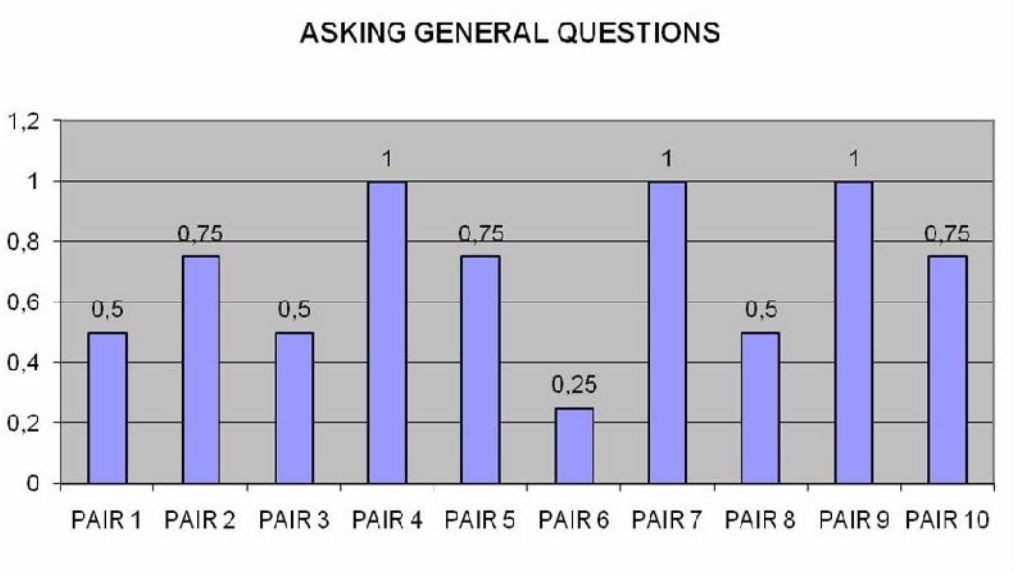
MEANS FOR PAIR 8

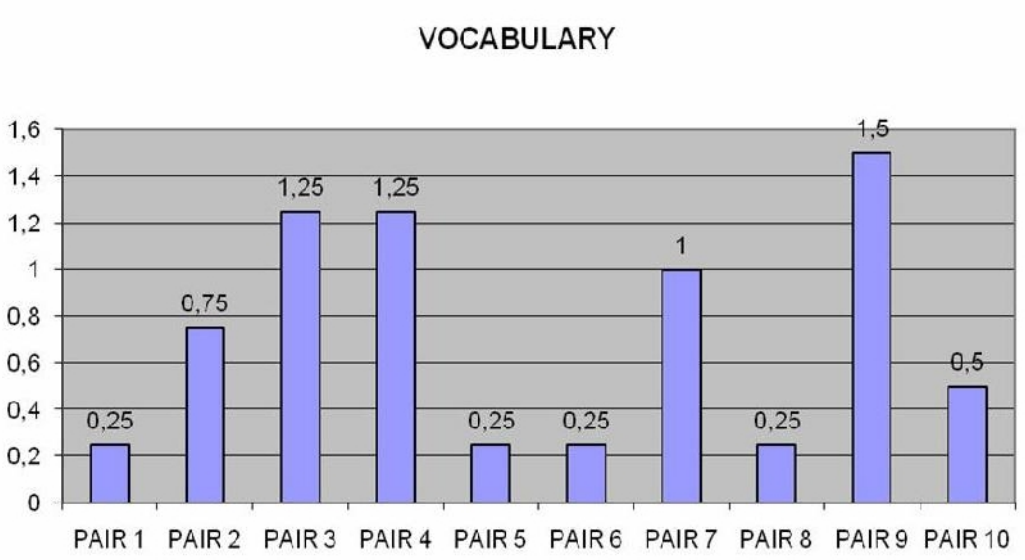


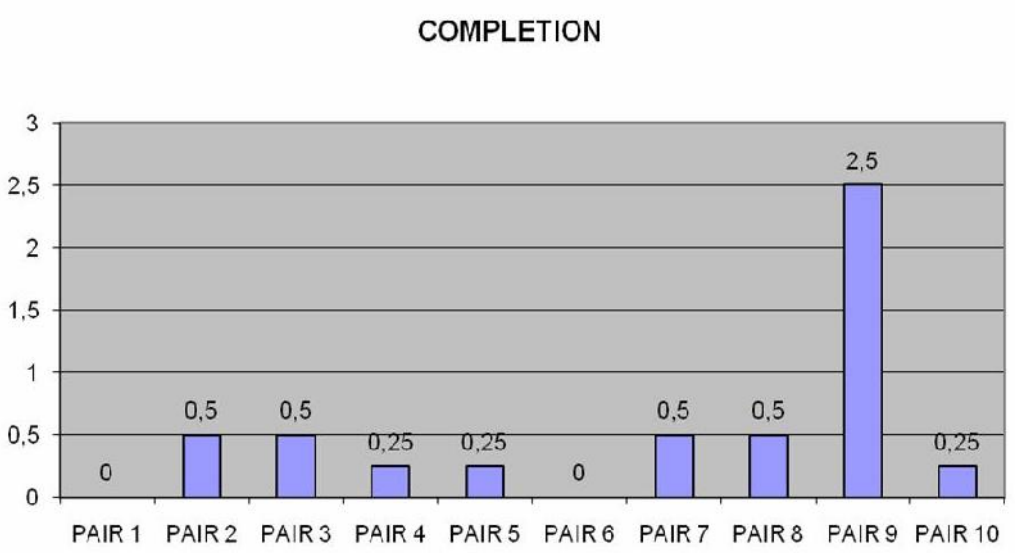

MEANS FOR PAIR 9
9
8
7
6
5
4
3
2
1
0
1
1,5
2,5
0,75
0,25
1,5
0,5
0,25
4,5
8
0
GENERAL QUESTION
VOCABULARY
COMPLETION
EXPLICIT RESPONSES
SPEED OF SPEECH
STRESS / PAUSE
CONFIRMATION STATEMENT
STATING CONCLUSION
CLARIFICATION
GIVING APPROPRIATE TIME
INTERRUPTING

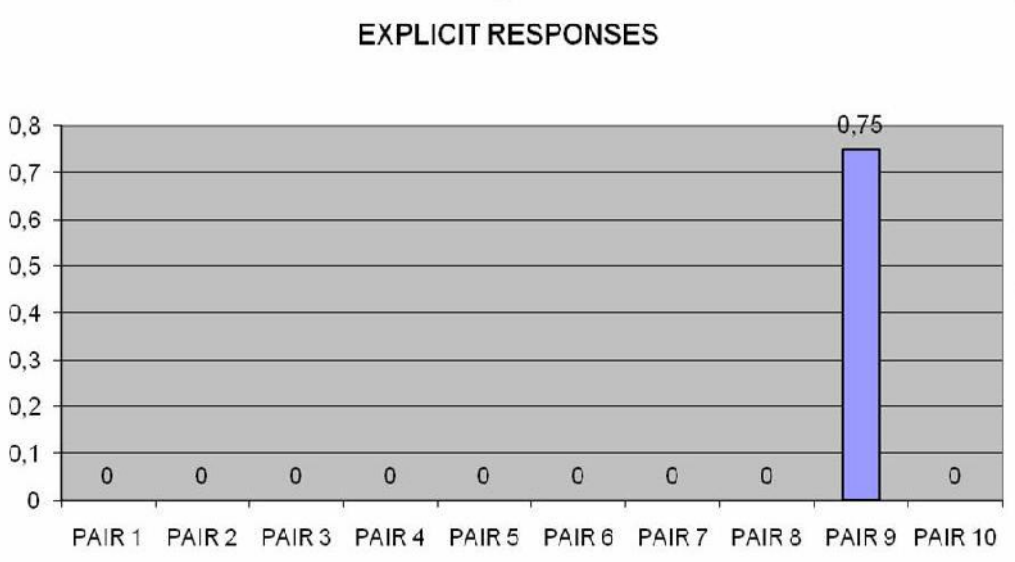

MEANS FOR PAIR 10
2,5
2
1,5
1
0,5
0
0,75
0,5
0,25
0
0
0,25
0
0,75
0,25
2,25
0
GENERAL QUESTION
VOCABULARY
COMPLETION
EXPLICIT RESPONSES
SPEED OF SPEECH
STRESS / PAUSE
CONFIRMATION STATEMENT
STATING CONCLUSION
CLARIFICATION
GIVING APPROPRIATE TIME
INTERRUPTING

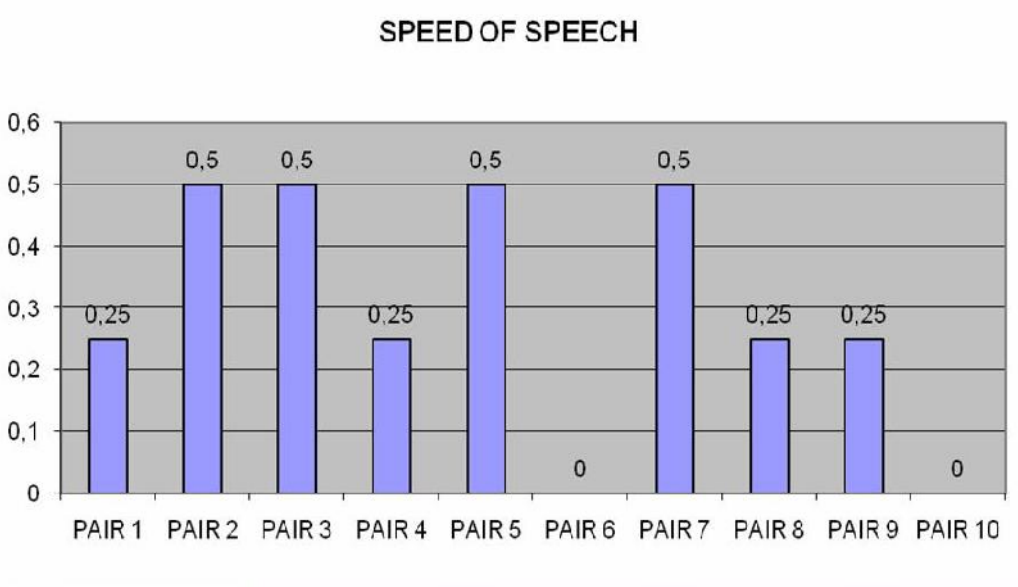
APPENDIX E: Figures of Means for Each Behavior in the Checklist

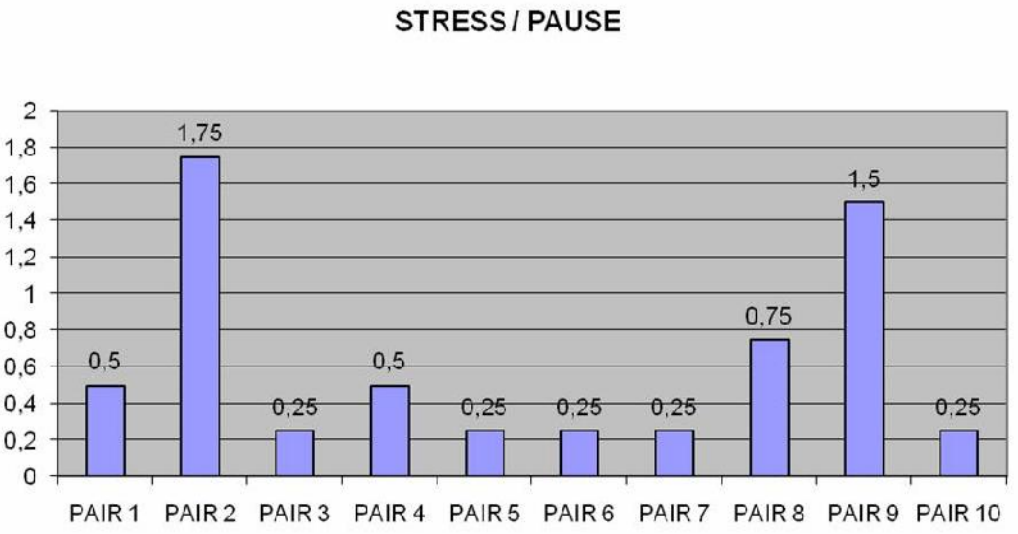


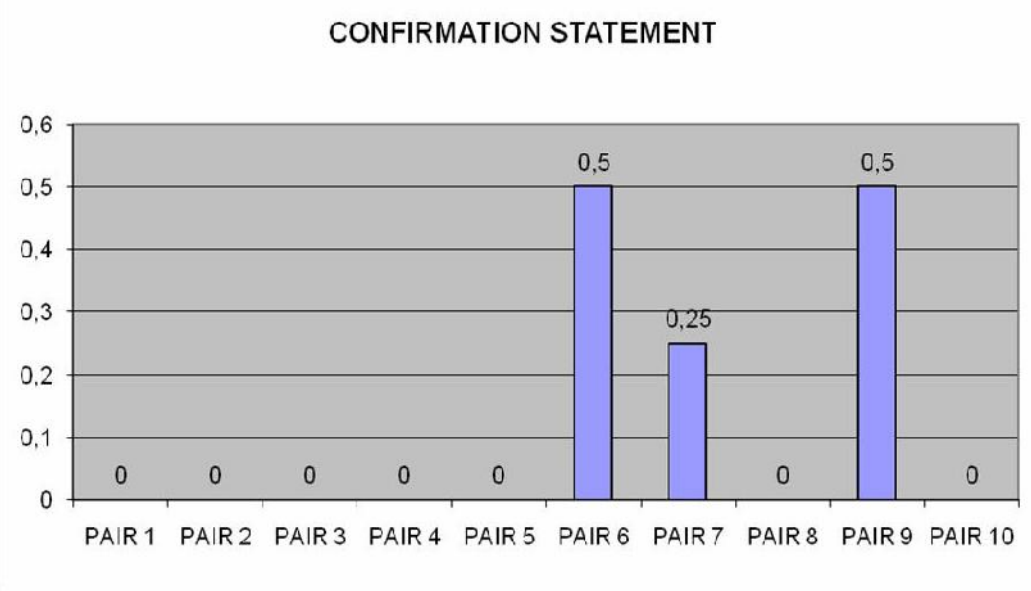


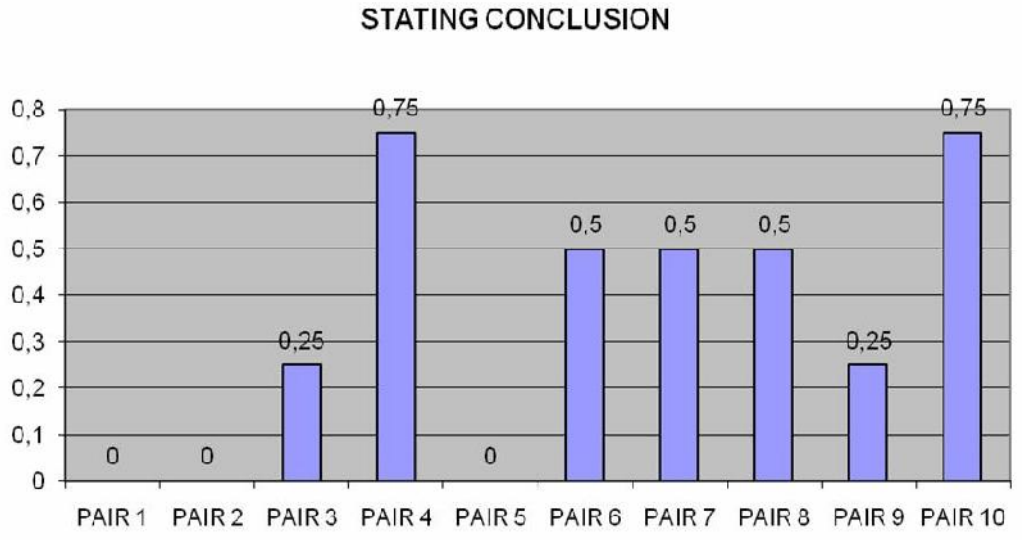

COMPLETION
3
2,5
2
1,5
1
0,5
0
0
0,5
0,5
0,25
0,25
0
0,5
0,5
2,5
0,25
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10

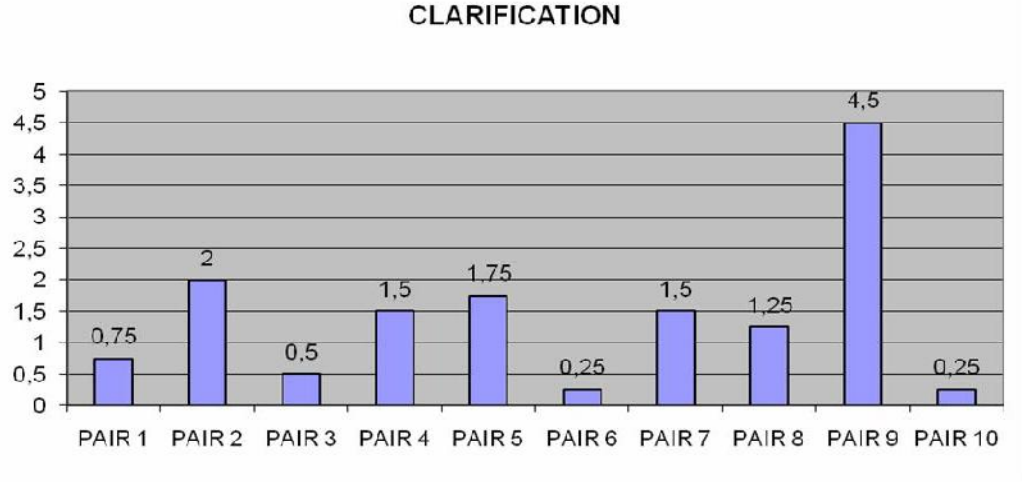

EXPLICIT RESPONSES
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
0
0
0
0
0
0
0
0,75
0
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10

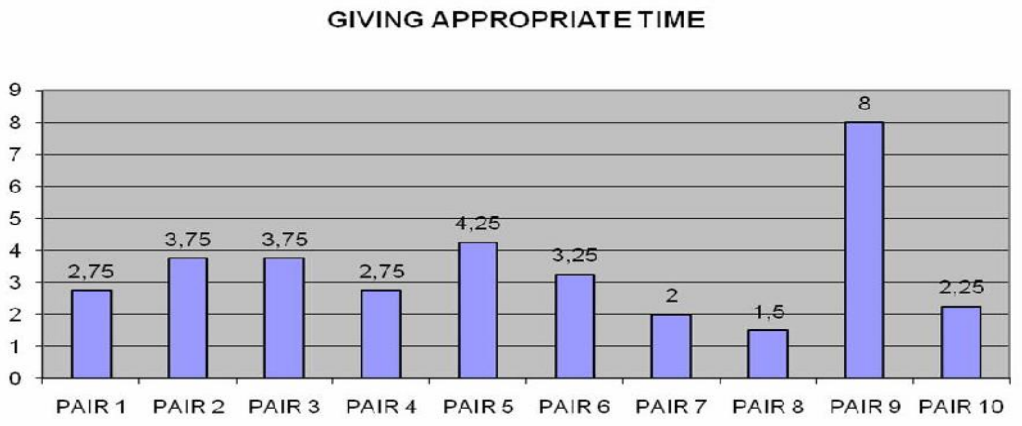

SPEED OF SPEECH
0,6
0,5
0,4
0,3
0,2
0,1
0
0,25
0,5
0,5
0,25
0,5
0
0,5
0,25
0,25
0
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10

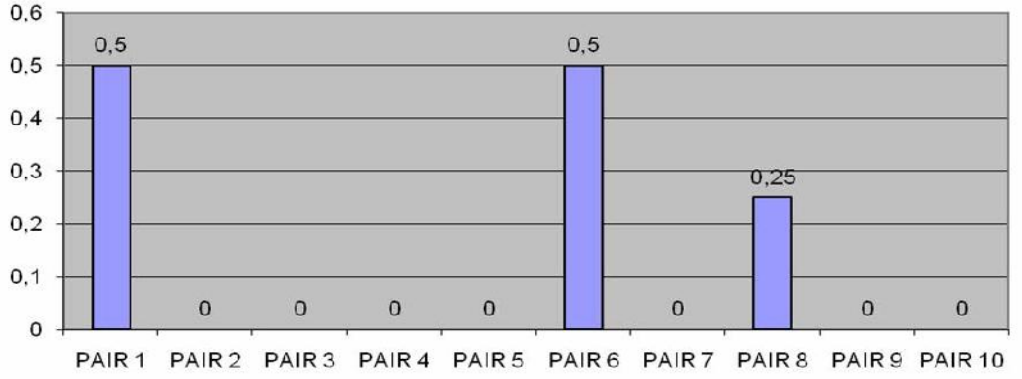

STRESS / PAUSE
2
1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
0
0,5
1,75
0,25
0,5
0,25
0,25
0,25
0,75
1,5
0,25
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10


CONFIRMATION STATEMENT
0,6
0,5
0,4
0,3
0,2
0,1
0
0
0
0
0
0
0,5
0,25
0
0,5
0
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10


STATING CONCLUSION
0,8
0,7
0,6
0,5
0,4
0,3
0,2
0,1
0
0
0
0,25
0,75
0
0,5
0,5
0,5
0,25
0,75
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10


CLARIFICATION
5
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
0,75
2
0,5
1,5
1,75
0,25
1,5
1,25
4,5
0,25
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10


GIVING APPROPRIATE TIME
9
8
7
6
5
4
3
2
1
0
2,75
3,75
3,75
2,75
4,25
3,25
2
1,5
8
2,25
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10


INTERRUPTING
0,6
0,5
0,4
0,3
0,2
0,1
0
0,5
0
0
0
0
0,5
0
0,25
0
0
PAIR 1
PAIR 2
PAIR 3
PAIR 4
PAIR 5
PAIR 6
PAIR 7
PAIR 8
PAIR 9
PAIR 10