To my beloved mother

A Comparison of Computer Assisted and Face-To-Face Speaking Assessment: Performance, Perceptions, Anxiety, and Computer Attitudes

The Graduate School of Education
of
Bilkent University

by

Ebru Öztekin

In Partial Fullfilment of the Requirements for the Degree of
Master of Arts

in

The Program of
Teaching English as a Foreign Language
Bilkent University
Ankara

July 2011

BİLKENT UNIVERSITY

THE GRADUATE SCHOOL OF EDUCATION

MA THESIS EXAMINATION RESULT FORM

July 18, 2011

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

Ebru Öztekin

has read the thesis of the student.

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

Thesis Title:	A Comparison of Computer Assisted and Face-To-Face Speaking Assessment:Performance, Perceptions, Anxiety, and Computer Attitudes
Thesis Advisor:	Asst. Prof. Dr. Phil Durrant Bilkent University, MA TEFL Program
Committee Members:	Asst. Prof. Dr. JoDee Walters Bilkent University, MA TEFL Program Assoc. Prof. Dr. Gölge Seferoğlu Middle East Technical University, Faculty of 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 Foreign Language.

_____

(Asst. Prof. Dr. Philip Durrant)

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 Foreign Language.

_____

(Vis.Asst. Prof. Dr. JoDee Walters)

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 Foreign Language.

_____

(Assoc. Prof. Dr. Gölge Seferoğlu)

Examining Committee Member

Approval of the Graduate School of Education

_____

(Visiting Prof. Dr. Margaret Sands)

Director

ABSTRACT

A COMPARISON OF COMPUTER ASSISTED AND FACE-TO-FACE SPEAKING ASSESSMENT: PERFORMANCE, PERCEPTIONS, ANXIETY, AND COMPUTER ATTITUDES

Ebru Öztekin

M.A. Department of Teaching English as a Foreign Language

Supervisor: Asst. Prof. Dr. Phil Durrant

July 2011

Computer technology has long been applied to language testing as a time and cost efficient way to conveniently assess the proficiency of large numbers of students. Thus, a good deal of research have focused on the effect and efficiency of computer assisted (semi-direct) assessment in evaluating different constructs of the language. Nonetheless, little research has been conducted to compare computer assisted and face-to-face (direct) formats to find whether the two modes yield similar results in oral assessment and whether one is advantageous over the other. Even less investigated were the possible outcomes of administration of computer-assessited speaking tests on a local basis, as achievement tests.

The purpose of this exploratory study is to fill the abovementioned gap via examining the relationships between a number of variables. Presented in the thesis are the relationships between test scores obtained in two different test modes at two different proficiency levels, the students' perceptions of the test modes, and their anxiety levels with regard to speaking in a foreign language, speaking tests, and

using computers. Data were collected through four computer assisted and four face-to-face speaking assessments, a questionnaire on Computer Asssisted Speaking Assesment (CASA) perceptions and another on Face-to-face Speaking Assessment (FTFsa) perceptions, a speaking test and speaking anxiety questionnaire, and a computer familiarity questionnaire. A total of 66 learners of English at tertiary level and four instructors of English participated in the study which was conducted at Uludağ University School of Foreign Languages.

The quantitative and qualitative data analyses revealed that the two test modes give very different rankings to the students, and the students' perceptions of the test modes, which have been found to be more positive about the FTFsa at both proficiency levels, are not strongly related to their performance in the speaking tests. The relationship between different types of anxiety mentioned above and test scores are only weakly related to the test scores and the degree of the relationships vary depending on the proficiency level.

The results of this study are hoped to be beneficial to the language assessors, instructors, and institutions and researchers that are into language assessment.

Key words: Computer assisted oral assessment, speaking assessment, face-to-face, speaking, speaking test, computer attitudes, anxiety

ÖZET

BİLGİSAYAR DESTEKLİ VE YÜZYÜZE YAPILAN KONUŞMA SINAVLARININ KARŞILAŞTIRMASI: PERFORMANS, ALGILAR, KAYGI, VE BİLGİSAYAR TUTUMLARI

Ebru Öztekin

Yüksek Lisans, Yabancı Dil Olarak İngilizce Öğretimi Bölümü

Tez yöneticisi: Yrd. Doç. Dr. Phil Durrant

Temmuz 2011

Bilgisayar teknolojisi uzun bir süredir zaman tasarrufu sağlayan ve düşük maliyetli bir yöntem olarak yabancı dil değerlendirmesinde kullanılmaktadır. Bu yüzden kayda değer miktarda araştırma dilin farklı yapılarını değerlendirmede bilgisayar destekli (yarı-dolaylı) değerlendirmenin etki ve etkinliğine yoğunlaşmıştır. Bununla birlikte, bilgisayar destekli ve yüzyüze formatların konuşma sınavında benzer sonuçlar verip vermediğini ve birinin diğerinden daha avantajlı olup olmadığını bulmak amacıyla az sayıda araştırma yapılmıştır. Bilgisayar destekli konuşma sınavlarının yerel düzeyde başarı sınavı olarak yapılmasının olası sonuçları ise daha az araştırılmıştır.

Bu keşif çalışmasının amacı yukarıda sözü edilen boşluğu bir dizi değişken arasındaki ilişkileri inceleyerek doldurmaktır. Bu tezde, iki farklı yeterlilik düzeyinde iki farklı sınav formatında yapılan sınavlarda alınan notlar, öğrencilerin sınav formatlarına yönelik algıları, ve yabancı dil konuşmaya, konuşma sınavlarına ve bilgisayar kullanımına dair kaygı düzeyleri arasındaki ilişkiler sunulmuştur. Veriler, dört adet bilgisayar destekli ve dört adet yüzyüze konuşma sınavı, bir Bilgisayar

Destekli Konuşma Sınavı (BDKS) Algıları ve bir Yüzyüze Konuşma Sınavı (YKS) Algıları anketi, bir konuşma sınavı ve konuşma kaygısı ölçeği, ve bir bilgisayar tutum ölçeği aracılığıyla toplanmıştır. Uludağ Üniversitesi Yabancı Diller Yüksekokulu'nda yürütülen çalışmada, toplamda İngilizce öğrenen 66 hazırlık sınıfı öğrencisi ve dört İngilizce öğretmeni yer almıştır.

Nicel ve nitel veri analizi, iki farklı sınav formatının öğrencileri çok farklı sıraladığını ve her iki yeterlilik düzeyinde de yüzyüze konuşma sınavı için daha olumlu olduğu bulunan sınav algılarının, öğrencilerin sınav performansıyla güçlü ilişkili olmadığını ortaya çıkarmıştır. Yukarıda söz edilen farklı kaygı türleri sınav notlarıyla zayıf ilişkilidir ve aralarında ilişkinin derecesi yeterlilik düzeyine göre değişmektedir.

Bu çalışmanın sonuçlarının yabancı dil değerlendirmesi yapanlar, yabancı dil öğretmenleri ve yabancı dil değerlendirmesiyle ilgilenen kurum ve araştırmacılara yardımcı olması umulur.

Anahtar kelimeler: Bilgisayar destekli sözlü değerlendirme, konuşma değerlendirmesi, yüzyüze, konuşma, konuşma sınavı, bilgisayar tutumları, kaygı, BDKS, YKS

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my thesis supervisor Asst. Prof. Dr. Philip Durrant, for his invaluable support, guidance, and feedback throughout the study. I also thank him for his encouraging and understanding attitude as well as his effort and assistance to better the thesis.

Many special thanks to Asst. Prof. Dr. Julie Mathews Aydınlı for her continuous support and Asst. Prof. Dr. JoDee Walters for her energy and enthusiasm as she motivated me to be an ever-progressing English teacher and broadened my horizons as an educator during the program.

I would like to show my gratitude to Saniye Meral, the associate director of Uludağ University School of Foreign Languages, and İsmet Çelik who made it possible for me to conduct my study in the institution.

I am also indebted to many of my colleagues for their support. This thesis wouldn't have been possible without the valuable contributions of Didem Gamze Dinç and Pelin Gümüş Sarıot who devotedly facilitated the flow of the study by dealing with the education and testing of the participants. I would like to thank to Rezan Altunbaş and Şeyda Özyürek who assisted with the scoring of participant performances. I am also grateful to Yakup Karakurt for the technical assistance he provided while installing and operating the software I used in the study.

I wish to thank to my dearest friend as a graduate student, Ebru Gaganuş, for helping me get through the difficult times, and for all the emotional support, entertainment, and caring she provided. Thanks to Ayça Özçınar, Zeynep Erşin, and

Bahar Tuncay for being a part of my life with the unbelievably nice moments we shared.

My greatest thanks to my mother Samiye Öztekin, my sister Seda Öztekin, and Ertuğrul Uzak, who are the meaning of my life, for their continuous encouragement and support during my hardest times throughout this program and my life.

Lastly, I offer my regards to all of those who supported me in any respect during the completion of my study.

TABLE OF CONTENTS

ABSTRACT iii

ÖZET v

ACKNOWLEDGEMENTS vii

TABLE OF CONTENTS ix

LIST OF TABLES xiv

LIST OF FIGURES xv

CHAPTER I: INTRODUCTION 1

Introduction 1

Background of the Study 2

Statement of the Problem 7

Significance of the Study 8

Research Questions 9

Conclusion 10

CHAPTER II: LITERATURE REVIEW 11

Introduction 11

Speaking as a Skill and Its Importance 12

Testing Speaking 14

Necessity and Different Ways 14

Qualities of a Useful Test 14

Computer Based Testing of Speaking 15

Advantages and Disadvantages of Direct and Semi-Direct Speaking Assessment....15

History and Examples of Computer Based and Other Semi-Direct Oral Tests....18

Validity, Reliability and Test Scores in Semi-Direct Tests of Oral Proficiency....22

Validity....22

Reliability....23

Perceptions of Test Takers....31

Anxiety and Performance in Oral Assessment....35

Anxiety Types and Definitions....35

The Relationship Between Speaking Test and Speaking Anxiety Level, and Performance in Oral Assessment....36

The Relationship Between Computer Attitutes and Performance in Oral Assessment....40

Conclusion....42

CHAPTER III: METHODOLOGY....44

Introduction....44

Setting....45

Participants....46

Instruments....47

Speaking tests....47

Questionnaires....56

CASA and FTFsa Perceptions Questionnaires 57
Speaking Test and Speaking Anxiety Questionnaire 57
Computer Attitudes Questionnaire 58
Data Collection Procedures 59
Data Analysis 60
Conclusion 62
CHAPTER IV: DATA ANALYSIS 63
Overview of the Study 63
Inter-Rater Reliability 65
A Comparison of the Scores 65
Were Average Scores Affected by Test Type, Level, Doing a Test in the First or Second place? 67
The Questionnaires 71
Item by item analysis of the CASA and the FTFsa perceptions Questionnaires 72
Qualitative Analysis of the Open-Ended Questions in the Perceptions Questionnaires 98
The Relationship Between Speaking Anxiety and Speaking Test Anxiety, Test Mode-Related Perceptions And Test Scores 102
The Relationship Between Computer Attitudes, Test-Mode-Related Perceptions And Test Scores 105
Conclusion 107
CHAPTER V: CONCLUSION 108

Introduction 108

Findings and Discussion 109

Performance, Reliability and Validity: The Scores Obtained in the CASA and the FTFsa 109

CASA and FTFsa Perceptions of the Test Takers 115

Speaking test and speaking anxiety and their relationship with the perceptions of the test modes and test scores 125

Computer attitudes and their relationship with the perceptions of the test modes and test scores 127

Pedagogical Implications 128

Limitations of the study 132

Suggestions for further research 134

Conclusion 135

REFERENCES 137

APPENDICES 143

APPENDIX A: Intermediate FTFsa 2 Questions 143

APPENDIX B: Intermediate CASA 2 Questions 150

APPENDIX C: Rating Scale 158

APPENDIX D: Band Descriptors 159

APPENDIX E: Expected Answers for the Second Intermediate CASA and FTFsa Test Questions 160

APPENDIX F: CASA Perceptions Questionnaire 162

APPENDIX G: Bilgisayar Destekli Konuşma Sınavlarıyla İlgili Tutum Ölçeği..... 166

APPENDIX H: FTFsa Perceptions Questionnaire 170

APPENDIX I: Yüzyüze Yapılan Konuşma Sınavlarıyla İlgili Tutum Ölçeği 174

APPENDIX J: Responses to the Open-Ended Questions in the CASA and the FTFsa Perceptions 178

Questionnaires (Turkish) 178

APPENDIX K: Foreign Language Speaking Test Anxiety and Speaking Anxiety Questionnaire 188

APPENDIX L: Yabancı Dil Konuşma Sınavı Kaygısı Ve Konuşma Kaygısı Ölçeği..... 193

APPENDIX M: Computer Attitudes Questionnaire..... 198

APPENDIX N: Bilgisayar Tutum Ölçeği..... 201

LIST OF TABLES

1 - Distribution of the Participants According to Levels and Groups 46
2 - Counter Balanced Design of the Study 55
3 - Inter-Rater Reliability Scores 65
4 - Comparison of Test Scores in the CASA and the FTFsa 66
5 - Reliability Analysis of the Perceptions Questionnaires 72
6 - Test-Mode-Related Anxiety Subscale (Overall) 74
7 - Test-Mode-Related Anxiety Subscale for Different Levels 75
8 - Questions about the Difficulty of the Speaking Tests (Overall) 81
9 - Questions about the Difficulty of the Speaking Tests as Perceived at Different Levels 82
10 - Questions about the Relationship Between Speaking Tests and Classroom Attendance 84
11 - Questions about the Relationship between Speaking Tests and Classroom Attendance at Different Levels 84
12 - Questions about the Quality of the Tests (overall) 85
13 - Questions about the Quality of the Tests at Different Levels 87
14 - Questions about the Comprehensiveness of the Tests 90
15 - Questions about the Comprehensiveness of the Tests at Different Levels 90
16 - Individual Questions that do not Belong to a Specific Category 92
17 - Responses to Individual Questions that do not Belong to a Specific Category at Different Levels 93
18 - The Relationships Between Speaking/Speaking Test Anxiety and Perceptions and Test Scores in Two Different Modes 103
19 - The Relationship between Computer Attitudes, Perceptions and Test Scores 106

LIST OF FIGURES

1 - Login page 49
2 - First page of the CASA 51
3 - Sample question in CASA 52
4 - Sample question 2 in CASA 53
5 - Administration sequence of the questionnaires 56
6 - The interaction between level and test mode 68
7 - The interaction between group and test mode 69
8 - The interaction between group, test mode, and level (pre-intermediate) 70
9 - The interaction between group, test mode, and level (intermediate) 70

CHAPTER I: INTRODUCTION

Introduction

A profound knowledge of language necessitates the mastery of various skills, the most challenging yet crucial one being oral language proficiency. There has been much controversy about whether speaking should be taught as a skill or only used as a means of teaching the language (Bygate, 2001), and so teaching speaking skills has been a thorny issue. However, once speaking is accepted as a skill of it is own and teaching speaking is stressed, it also becomes necessary to assess it (Larson, 2000). Assessing oral skills, which embody a complex range of hard-to-measure subskills, is no less complicated than teaching speaking. An array of factors such as the reliability, validity, and fairness of the test, as well as the accuracy, consistency, and ability-representativeness of the instruments should be considered before administering a speaking test.

Despite the difficulty of testing oral skills, a variety of ways, most of which are of a face-to-face format, have been successfully used to assess the speaking ability of the learners over a long time. Nevertheless, due to the labor extensive, costly and time-consuming nature of face-to-face speaking tests, some schools or teachers feel obliged to abandon the task of testing speaking skills or simply tend to ignore the need to assess it. This may influence, although indirectly, learners' motivation for improving their speaking skills. As a result, even learners with high proficiency with regard to the knowledge of structures or receptive skills of the language may fail to perform successfully when it comes to speaking.

A practical, time and cost efficient solution for the schools that have difficulty in investing time and expertise in the assessment of oral skills may be the use of computer technology. Effective implementation of computer technology in assessing skills other than speaking is now a well-known phenomenon in the language testing world. It has been widely used to assess grammar, vocabulary, reading, listening and writing skills by international testing organizations in particular. The use of computer based assessment of speaking skills, on other hand, is a less widely-used yet promising approach. The present study will investigate the possibility of using semi-direct, in other words, computer-mediated, oral assessment in an attempt to offer a feasible alternative to face-to-face assessment of oral skills at local schools, which might be an ideal method to avoid the drawbacks of the latter.

Background of the Study

Recently there has been growing interest in the utilization of technology in language education as a supplement to conventional lessons. In line with its usage within the classroom and materials development, computer technology has also been used for language testing as a time and cost efficient way to conveniently assess the proficiency of large numbers of students. As Chapelle and Douglas (2006) note, technology has been increasingly applied to almost all aspects of language testing, including test development, test delivery, and rating. Although computers have been widely used for the assessment of many skills and structures, using them to facilitate oral assessment in particular can be considered a relatively unexplored area.

Luoma (2004) states that the face-to-face mode is the most common way of assessing oral proficiency. However, this does not necessarily mean that it has been

the only way ever used. As early as in 1979, Clark distinguished between three types of speaking tests, namely, indirect, direct and semi-direct (O'Loughlin, 2001). Clark defines indirect tests as procedures where the test taker is not actually required to speak. Direct tests are procedures in which the examinee is engaged in face-to-face interaction with one or more interlocutors, whereas semi-direct tests elicit active speech from the test taker by means of tape-recordings, printed test booklets, or other nonhuman elicitation procedures. The semi-direct test format was later named the Simulated Oral Proficiency Interview (SOPI; Stansfield and Kenyon, 1992), a similar name to that of its direct counterpart: Oral Proficiency Interview (OPI). The fact that the term "semi-direct" evolved in the 1970s shows that educators have actually sought alternative ways to assess orals skills for a long time, especially to be able to assess the oral proficiency of large numbers of students with ease. Originally, semi-direct testing was used to refer to a tape-recorded procedure accompanied with a test booklet, which makes sense when the technology of the 1970s is taken into account. However, today, the term is also used interchangeably with Computer-Assisted Assessment, Computer-Aided Assessment, Computer-Mediated Assessment, or Computer-Based Assessment of speaking skills (Douglas and Hegelheimer, 2007; Galaczi, 2010; Winke and Fei Fei, 2008) because computers provide the latest technology for assessment. Despite the slight differences between the terms listed above, most researchers use them to refer to the same format of speaking tests: the semi-direct mode. Using semi-direct, or Computer-Assisted Speaking Assessment (CASA), in delivering and administering tests now appears to be a popular practice for professional testing organizations. Instances include the oral subtest of the internet based Test of English as a Foreign Language (TOEFL IBT) of the

Educational Testing Service in the United States , and the Graduating Students' Language Proficiency Assessment–English (GSLPA) speaking component developed by Hong Kong Polytechnic University.

Previous studies on semi-direct testing have mostly focused on considering the advantages and disadvantages of semi-direct assessment. Qian (2009) lists some advantages of semi-direct assessment over face-to-face oral assessment. First, semi-direct oral assessment economizes on the expert resource, as the expert is not obliged to be on-site during the examination. Secondly, it is both cost-effective and efficient: a single version of the semi-direct test can be administered to a large number of test takers at the same time or within a very short period. O'Loughlin (2001) also states that semi-direct tests represent a more standardized and cost efficient approach to the assessment of oral language proficiency than their direct counter-parts. Another advantage, pointed out by both Qian and O'Loughlin, concerns test reliability and fairness as the test taker will receive standardized instructions and prompts.

On the other hand, semi-direct assessment is supposed to be inferior to direct testing in terms of test validity as real-life communication typically takes place face-to-face (van Lier, as cited in Qian, 2009). The language produced through the semi-direct mode is considered artificial as the test taker has to speak into a recorder to a disembodied interlocutor. Citing the growing evidence gathered which favors direct over semi-direct in terms of validity, Cheng (2008) states that it is not certain whether semi-direct tests can replace direct ones. O'Loughlin (2001) investigated the construct validity of the two test formats and whether they can be considered equivalent in theoretical and practical terms by examining the oral component of the *access:* test (the Australian Assessment of Communicative English Skills). The

conclusion O'Loughlin arrives at is that the spoken interaction of two or more people is jointly constructed and hence fundamentally different in its character from communication with a machine. The researcher therefore cautions against using the direct and semi-direct forms of the test interchangeably. Shohamy (1994) also suggests that direct tests and semi-direct tests measure different constructs, which means that a semi-direct test is prone to lack construct validity if it attempts to measure the same construct as a direct test. Bailey (2006) also claims that, although indirect tests are highly practical, their face validity is always in question.

Other studies into direct and semi-direct testing have indicated that there is considerable overlap between direct and semi-direct tests in terms of the skills they tap, at least in the sense that people who score high in one mode also score high in the other (Luoma, 2004). In support of this view, Qian (2009) reports that research based on concurrent validity has provided statistical evidence (r = 0.89–0.95) that the direct and semi-direct testing modes of the same test produces comparable scores. Shohamy (1994) also reports that the concurrent validity of the two types of tests is high. Xiong, Chen, Liu, and Huang (as cited in Cheng, 2008) reported a high correlation between students' ranking in class and their scores from the semi-direct speaking test , which led to the interpretation that the students demonstrated their actual oral language proficiency through the semi-direct test. As a result, the semi-direct test was deemed to be a feasible alternative to direct test by the researchers.

Another focus of interest has been the question of whether the conditions of semi-direct, or computer mediated tests have any influence on the degree of anxiety that students experience. One claim is that the anxiety levels of the test takers differ since they feel more nervous about the test because everything they say is recorded

and no gestures or expressions can be used which leaves speaking as the only channel. The findings of Guo (as cited in Cheng, 2008) support this view. Guo tested ten final year English majors in three situations: (a) recording their opinions of a topic on a tape (b) talking to some freshmen in a casual environment; and (c) talking to a tester in an office. The researcher concluded that the pressure felt by students in different situations led to various degrees of anxiety which affected their fluency. The participants were most fluent and least anxious in the casual environment, whereas the tape-based version (first situation) caused more anxiety. Although the generalizability of the research is hindered by the limited number of participants, the researchers' suggestion of considering test takers' affective factors while developing oral tests is worth investigating.

Briefly, there are conflicting results and ongoing discussions on the necessity, validity and efficiency of semi-direct oral assessment and its equivalence to its more conventional counterpart, face-to-face, or direct oral assessment. Qian (2009) suggests that because of the increasing popularity of semi-direct, or, in more contemporary terms, computer-based oral language assessment, there is a need to evaluate further the potential merits and problems as associated with it.

Although considerable research has been devoted to the use of semi-direct, or computer mediated oral assessment as counterparts of widely applied direct oral examinations by well-known institutions, little attention has been paid to its use by smaller institutions such as preparatory schools or colleges as part of their regular assessment procedures. Luoma (2004) claims that in a practical sense, the sheer amount of work required for developing a tape-based (semi-direct) test makes it impractical for classroom testing. However, with the new advances in computer

technology, it now seems possible to produce computer mediated oral assessments and have them rated by expert humans later on. The main purpose of the study reported here is, thus, to investigate the advantages and disadvantages of relatively small scale computer mediated oral assessment applied in local institutions by comparing the two modes. The present work also differs from previous studies by investigating students' performance and anxiety level in both modes taking their language proficiency levels as well as their computer attitudes into account. Finally, the experiment includes an exploration of test takers' perceptions regarding the two modes addressed.

Statement of the Problem

There has been a considerable focus on the use of computer technology in language assessment (Chapelle and Douglas, 2006). Nevertheless, using computers as mediators for oral assessment is a relatively new area of interest to researchers. Therefore, potential merits and problems as associated with computer based oral assessment need to be further evaluated (Qian, 2009). Some of the literature in this area suggests that computer mediated speaking tests, or semi-direct tests, can not replace face-to-face, or direct tests of oral proficiency. Underhill (1979) is strongly critical of the lack of authenticity of semi-direct direct tests, whereas other researchers (i.e. O'Loughlin, 2001) criticize the construct validity of these tests. On the other hand, features such as practicality, cost-effectiveness and reliability may make semi-direct tests a feasible alternative for face-to-face oral assessment. The ongoing debates among researchers as to whether computer mediated oral assessment is equivalent to traditional face-to-face format, and whether it is even more advantageous to use than the former remains inconclusive. Therefore, in-depth

analysis of the effects and efficiency of computer-based oral assessment is essential to be able to decide whether it can safely replace the face-to-face format.

Tertiary schools responsible for language education in the EFL setting of Turkey have problems in assessing the oral skills of large numbers of students. Moreover, institutions conducting nation-wide language examinations lack a component assessing oral proficiency, thus a solution for larger scale examinations is also a need. When the apparent lack of appropriate speaking tests is considered, analyzing the pros and cons of computer mediated or semi-direct oral assessment by referring to affective effects on test takers and features of the semi-direct speaking tests as well as how they are perceived by the students may illuminate the way to preparation of successful computer mediated oral assessments.

Significance of the Study

The literature on different modes of oral proficiency assessment has offered contradictory findings about the appropriateness of semi-direct testing with respect to its validity and equivalence to the face-to-face format which is presumed to be the best way of assessing oral proficiency (Luoma, 2004). The present study is intended to reflect on the use of computer mediated oral assessment as a substitute for a face-to-face format at a Turkish tertiary school at pre-intermediate and intermediate levels. It aims to contribute to the current literature by shedding light on the degree to which computer mediated oral assessment is valid and equal to the face-to-face format when used at tertiary level in EFL settings. The findings of this study can strengthen an argument for or against the use of semi-direct tests in oral language proficiency testing and provide researchers as well as educators and administrators

with up-to-date information regarding the issue as little research has been conducted to investigate the effects of the latest technology on oral assessment.

At the local level, the study is expected to provide administrators and teachers with up- to-date information on a standardized, time and cost efficient way of conducting oral assessment which arguably has higher reliability in comparison with the face-to-face format as the latter is basically reliant on subjective scoring of raters on-site. Information gathered on the usability of computer mediated oral assessment may be valuable to preparatory schools of universities in Turkey (such as Uludağ University School of Foreign Languages – the setting where the actual investigation will take place) which plan to make use of new technologies in teaching and assessment. In a broader sense, the application itself can set a precedent for the nation-wide language proficiency tests, none of which currently features an oral language component due to shortcomings in expertise and financing.

Research Questions

The research questions to be addressed in the study are:

1. How is the speaking performance of the pre-intermediate and intermediate level test takers at tertiary school affected by the test mode being either the face-to-face (FTFsa) or the computer-assisted speaking assessment (CASA)?
2. What are the test takers' perceptions of oral assessment?
    a. What are their perceptions of the FTFsa?
    b. What are their perceptions of the CASA?
3. What is the relationship between the anxiety levels of the test takers and the test mode?

   a. What is the relationship between speaking/speaking test anxiety, and test scores?
   b. What is the relationship between speaking/speaking test anxiety, and students' perceptions of FTFsa and CASA?
4. What is the relationship between the computer attitudes of the test takers and the test mode?
   a. What is the relationship between students' computer attitudes and test scores?
   b. What is the relationship between students' computer attitudes and their perceptions of FTFsa or CASA?
5. Depending on the test mode, do the speaking performances, test-mode-related perceptions, anxiety levels of the test takers at different proficiency levels differ?

Conclusion

In this chapter, the background of the study, statement of the problem, significance of the study, and the research questions were presented. The next chapter will review the relevant literature. In the third chapter, the methodology including the setting and the participants, instruments, data collections methods and procedures will be described. The data collected will be analyzed and reported quantitatively and qualitatively in the fourth chapter. Finally, the fifth chapter will present the discussion of the findings, pedagogical implications, limitations of the study, and suggestions for further research.

CHAPTER II: LITERATURE REVIEW

Introduction

This empirical study investigated the advantages and disadvantages of direct and semi-direct forms of speaking assessment through the evaluation of a face-to-face speaking assessment (FTFsa) and a specifically developed computer-assisted speaking assessment tool (CASA). It draws on data from the tests themselves and from three questionnaires to examine the scores obtained, student perceptions and levels of various types of anxiety experienced in both modes at two proficiency levels: pre-intermediate and intermediate. The data were mainly analyzed using quantitative methods, supported by qualitative analysis of some information from the questionnaires. Pursuant to the analyses, suggestions regarding the choice of speaking test mode were made, which were supposed to be of assistance to the teachers of English and the administrators as well as the EFL learners.

This chapter consists of multiple sections. The first section reviews the literature on the definition and importance of the speaking ability in English language teaching. This is followed by the second section on the necessity of the assessment of speaking ability and types of assessment as well as qualities of a good speaking test. The third section provides an insight into the attributes of computer assisted speaking assessment and its history. Fourth comes the section focusing on the literature on the validity and reliability of the semi-direct assessment of oral proficiency. It is followed by the fifth section about test takers' perceptions of the test modes, the sixth, which is about the relationship between two different types of

anxiety and test performance, and finally the seventh about the relationship between computer attitudes and test performance.

Speaking as a Skill and Its Importance

Chaney and Burk (1998) define speaking as the process of building and sharing meaning through the use of verbal and non-verbal symbols, in a variety of contexts. Speaking in a foreign language is not very easy and it usually takes a long time to become competent. This may be due to the fact that the speaking skill comprises a number of other macro and microskills which constitute the whole skill when brought together. Brown (2004) defines these microskills as the skills of producing the smaller chunks of language such as phonemes, morphemes, words, collocations, and phrasal units. The macroskills refer to larger elements such as fluency, discourse, function, style, cohesion, nonverbal communication and strategic options. It is not vital for learners to have metalinguistic awareness of the components of the speaking skills in order to use them effectively (Bailey, 2006). Yet, learners are expected to be able to learn and use these components since, as Bailey suggests, speaking might be accepted as the most fundamental of human skills. Moreover, speaking has been recognized as an interactive, social and contextualized communicative event. Therefore, it has a key role on developing students' communicative competence (Usó-Juan & Martínez-Flor, 2006). Given that speaking proficiency is one of the basic constituents underlying communicative competence, it is obvious that teaching speaking is an important part of second language education.

In spite of the fact that speaking is now valued by language educators, this was not the case a few decades ago. As Bygate (2001) notes, only recently has speaking started to emerge as a separate skill to be taught or tested. Bygate proposes three reasons for this, the first being *tradition,* which refers to the considerable effect of grammar translation approaches on language teaching. The second factor is *technology:* the equipment, i.e. tape- recorders or computers, required to study speaking through hearing speech samples was not adequately available until recently, which led to a focus on the written rather than spoken form of the language. The third factor delaying the perception of speaking as a skill of its own is *exploitation.* Most approaches, including the direct method, Community Language Learning and the Silent Way, recognized oral communication merely as a special medium for providing language input, memorization practice and habit formation; it was not taught as a discourse skill in its own right. The Audio-lingual Method (ALM) was one of the first approaches focusing on the teaching of oral skills. Nevertheless, teaching of oral language was limited to engineering the repeated oral production of structures in the target language (Bygate, 2001). That is, oral language was only a medium in ALM as well. As Bygate mentions, upon realizing that ALM neglected the relationship between language and meaning in addition to the importance of social context, two types of communicative approach, namely, a notional-functional and a learner centered approach, were developed around the 1970s. The former attempted to include interactional notions in grammar teaching, and the latter concentrated on meeting the expectations of learners in terms of communicating meaning. The latest trend in the teaching of speaking skills is the task-based

approach where skills-based models have been used. Briefly, speaking has its own place in language teaching now.

Testing Speaking

Necessity and Different Ways

According to researchers focusing on the assessment of oral proficiency (Larson, 2000; Luoma, 2004), the fact that speaking skills are an important part of the curriculum in language teaching makes them an important object of assessment as well. This has led researchers to seek feasible, efficient and practical tasks, criteria and modes (or formats) for assessing oral proficiency. Among numerous task types for assessing speaking, Thornbury (2005) identifies interviews, live monologues, recorded monologues, role-plays, collaborative tasks and discussions. Instances of commonly used criteria to assess speaking are the American Council on the Teaching of Foreign Languages (ACTFL) Speaking Scale, the Common European Framework (CEF), and the Test of Spoken English (TSE) band descriptors by ETS. Finally, the formats, or modes, of speaking assessment as defined by Clark in 1979 are direct, indirect and semi-direct modes of oral assessment, which are the focus of the present study.

Regardless of the mode they are administered, the tests should have certain qualities to be considered useful tests. Therefore, the attributes that semi-direct tests, as well as direct and indirect tests, should bear are presented below.

Qualities of a Useful Test

The most important quality of a test is its usefulness, that is, whether it serves the purposes it is intended for (Bachman & Palmer, 1996, p. 17). Bachman and

Palmer identify a test usefulness model consisting of six test qualities: reliability, construct validity, authenticity, interactiveness, impact and practicality, suggesting that there should be an appropriate balance among these qualities, since different combinations of them affect the overall usefulness of a particular test. Similarly, McNamara (2008) states three basic dimensions - validity, reliability and feasibility - the needs of which should be balanced depending on the text context and test purpose. Discussions by researchers on qualities such as interactiveness, practically and feasibility of semi-direct speaking assessment are presented in the next section. Validity and reliability will be defined later in this chapter along with reports of the empricical studies that sought for the validity and reliability of the semi-direct tests of speaking ability because these two are among the most commonly investigated qualities of the semi-direct speaking tests in the relevant literature.

Computer Based Testing of Speaking

Advantages and Disadvantages of Direct and Semi-Direct Speaking Assessment

As for speaking tests, the application of which is becoming more desirable each day with the increasing importance given to speaking proficiency, McNamara proposes that feasibility can only be achieved through semi-direct tests. He adds that the semi-direct format is practical as it can be administered on demand in any location, fair because the interlocutor effect is eliminated - all candidates receive the same prompts -, and economical since there is no need for an on-site interlocutor (McNamara, 2008) and, as Qian (2009) suggests, a single version of the test can be administered to large numbers of test-takers, which economizes on test development resources. In addition, since the responses are recorded, the marking process can take

place anytime and anywhere. Throughout the marking process, the raters can simply skip the instruction parts and listen to the answer of the test taker, saving time. The fact that the candidate output content is predictable facilitates the construction of accurate scoring criteria, which is said to yield more reliable results (Underhill, as cited in O'Loughlin, 1997). In addition, as the use of semi-direct tests would increase the number of students who have a chance of taking speaking tests, students and educators will probably invest more in developing second language speaking skills. This potential for positive washback is especially important in settings where the oral proficiency of huge numbers of students should be assessed but it is impossible to do so due to practical concerns (Yu & Lowe, 2009).Taking the recent developments regarding language portfolios into account (i.e. Chang, Wu & Ku, 2005), it is also possible to suggest that voice recordings captured via semi-direct tests can be used as a part of the candidate's portfolio, demonstrating the improvement in his speaking ability over time and proving his final speaking proficiency (Huang & Hung 2010).

A study by Larson (2000) seems to support the view that semi-direct speaking tests are advantageous. Larson mentions the use of a computer program for oral assessment, Oral Testing Software (OTS), developed by Brigham Young University (BYU), and reports the results obtained via piloting the software by conducting achievement tests to see BYU students' progress in oral language competency. Initially, audio cassette players were used at BYU for oral assessment due to the need to test oral skills on a frequent basis in a limited time. However, it was discovered that scoring the test tapes still required a substantial amount of teacher time as teachers lost time while listening to the sections consisting of warm-up questions before hearing the answers to the actual test questions. As a result, the computerized

version, the OTS, was introduced. Larson lists numerous advantages of the computerized speaking test over tape-mediated and face-to-face forms. First, due to the enhanced quality of voice recordings, it became easier for the raters to discriminate between the sounds heard and to rate them fairly. As compared with the face-to-face form, in the OTS, all testees received an identical test, which means that they received the same questions in exactly the same way within the same time limits. In addition, they did not have the chance to manipulate the examiner to their advantage. As compared with the tape-mediated form, it was possible to use a wider range of prompts (visuals, audio-visuals, graphics and texts) to elicit the answer in the OTS. The access to student responses to evaluate them was also facilitated. Finally, Larson reports that only minimal computer literacy was adequate both for the teachers to administer the test and for the students to take it.

On the other hand, semi-direct tests of oral proficiency also have their inherent drawbacks. O'Loughlin (1997) asserts that semi-direct speaking tests usually elicit speech in the form of monologues. He further claims that monologic talk is more difficult than conversations for some language learners. Moreover, Clark (1979) notes that these tests are less real life like, and thus, can only be second order substitutes for live interviews. In other words, they cannot be used instead of face-to-face speaking tests at all times as the two modes are not equivalent, yet, using a semi-direct mode is still an option.

Direct tests of speaking proficiency, namely, live interviews, are also problematic in many ways, however. Hughes (1989) argues that the relationship between the interlocutor and the test taker is asymmetrical, that is, the latter is usually unwilling to take the initiative and start the conversation. As a result, some

styles of speech, such as asking for information, can rarely be elicited in direct tests. Considering the fact that there is an interlocutor who tries on purpose to speak to the examinee and make him speak, the interviews are also said not to be real life like (Clark, 1979). There are also some problems related to the raters. For instance, Luoma (2004) reports that the lack of anonymity in face-to-face speaking tests and the fact that different raters attend to different aspects of the speech yields unreliable results. Similarly, McNamara (2008) states that some raters might be lenient to some types of errors, might tend to focus more on grammar, or differ in interpreting the rating scale, which would result in low reliability.

With all the advantages of the semi-direct format and the disadvantages of the direct format considered, as a useful testing format, the semi-direct oral assessment may be nominated as a reasonable alternative to the direct, face-to-face mode of oral assessment, or they may be combined to eliminate the disadvantages of either test mode. Numerous researchers conducted studies or developed speaking tests with the aim of successfully implementing the semi-direct speaking tests as substitutes for direct ones. Presented below are some examples from the earlier or existing semi-direct speaking tests to provide a better insight into where and how these tests can be utilized.

History and Examples of Computer Based and Other Semi-Direct Oral Tests

Numerous semi-direct tests have been developed in an attempt to find alternative ways to evaluate the second language speaking ability of large numbers of students in a practical way. TSE (Test of Spoken English), one of the earliest examples of such tests, was developed by Clark and Swinton in 1979 as a part of the renowned TOEFL (Test of English as a Foreign Language) to complement its

listening and reading components. Other examples include the MLA Cooperative speaking tests and the tape and booklet-mediated speaking tests of the ETS Advanced Placement Program. In 1980, Rowe and Clifford developed the ROPE (Recorded Oral Proficiency Examination) consisting of tape-recorded questions, replies to which were recorded on tapes by the examinees. As Clark (1986) notes, the ROPE had been the only example of "proficiency oriented semi-direct tests" until 1984, when Clark started a project with the aim of developing a tape-based test of Chinese speaking proficiency. This test differed from the ROPE in that in addition to the audio-tape, it also included a printed test booklet, which consisted of visuals and text contributing to the meaning of the questions heard on the tape. Another version of the semi-direct tests was created and improved through the joint efforts of the language assessors Clark and Li in 1986 at the Center for Applied Linguistics (Stansfield, 1990). This test was later titled the Simulated Oral Proficiency Interview, or SOPI (Stansfield & Kenyon, 1988), which later became to be utilized around the world.

Kenyon and Malone (2010) provide a list of SOPIs that were developed in other languages after the Chinese version: Portuguese (Stansfield, Kenyon, Paiva, Doyle, Ulsh & Cowles, 1990), Hausa (Stansfield & Kenyon, 1993), and Indonesian (Stansfield & Kenyon, 1992). In the 1990s, the Chinese Speaking Test was updated and new tests in Russian, Spanish, French, and German were generated. The main reason behind the creation of the SOPI was the necessity to find a way of using the common ACTFL OPI speaking proficiency guidelines for less commonly taught languages, which was a challenge due to the limited number of trained interviewers. As Kenyon and Malone report, a more developed version of the SOPI, COPI

(Computerized Oral Proficiency Interview), was developed during the same decade. The COPI was designed as an adaptive test during which test takers are given the chance to choose from a range of topics and difficulty levels to demonstrate their existing proficiency. The test takers can also control the planning and response times to some extent as well as the instruction language, be it in their mother tongue or second language. Compared with the OPI, the SOPI/COPI are disadvantageous in one respect: the prompt is one-way in the SOPI/COPI whereas there is a two-way conversation in the OPI; that is, the examinee can request clarification, repetition, or restatement, and the interviewer can modify the conversation accordingly (Kenyon & Malone, 2010).

Another instance of semi-direct tests is the PPS ORALS, (the Pittsburgh Public Schools Oral Ratings Assessment for Language Students), a grant-funded project to create online testing software that makes district-wide oral testing feasible. The PPS ORALS assessment model is proposed as a valid instrument for determining students' oral proficiency in accordance with the ACTFL Oral Proficiency Scale. The PPS ORALS project proved to be a valid, reliable and feasible performance based assessment of oral proficiency after four years of trial (Fall & Glisan, 2007).

The English Test of the Graduating Students' Language Proficiency Assessment (GSLPA), first implemented at the Hong Kong Polytechnic University (HKPU) in the 1999/2000 academic year, consists of writing and speaking sections. Conducted at multimedia language libraries in 40 minutes as an exit test for the university in semi-direct format, the speaking component has five tasks:

> "Summarizing and reporting information from a radio, responding to a series of questions at a job interview, presenting information from a written (graphic) source to a business meeting, leaving a work-related telephone message, providing information about an aspect of life in Hong Kong to a newly-arrived international colleague" (Qian, 2007).

A very well-known instance of computerized assessment of oral proficiency is the speaking component of the Test of English as a Foreign Language™ Internet-based test (TOEFL® iBT Speaking test) of the Educational Testing Service (ETS), first introduced in 2005. The speaking test is composed of six tasks – two independent and four integrated tasks - requiring test takers to wear headphones and speak into a microphone as they respond. The responses are recorded digitally and rated by certified ETS raters (Xi, 2008).

Another example of validated computerized or tape-based semi direct tests of oral proficiency is PhonePass SET 10 (Bernstein, De Jong, Pisoni & Townshend, 2000), a test administered over the telephone via a computer system. The difference of PhonePass from other semi-direct oral assessment instances is its fully automated nature, where the scoring is also done by the computer system.

Except for Larson (2000), all of the widely known taped or computerized semi-direct tests of oral proficiency mentioned above are tests used nation-wide or internationally with the purpose of assessing examinees' overall speaking ability. In other words, they are proficiency tests questioning how much global competence one has in a language, as defined by Brown (2004). There has been little focus on computer assisted assessment of oral skills in the form of achievement tests, which are directly related to classroom lessons or the total curriculum (Brown, 2004) and are typically used at the end of a period learning (Davies, Brown, Elder, Hill, Lumley & McNamara, 2002).

Validity, Reliability and Test Scores in Semi-Direct Tests of Oral Proficiency

Validity

One of the crucial qualities sought for in a test is validity. As Hughes (2000) states, a test is considered valid if it measures what it is supposed to measure. Luoma (2004) asserts that validity refers to the meaningfulness of scores. The concept of validity comprises a number of aspects, though, and there are different types of validity that address different aspects. Among the aspects to be touched upon in this study are content, construct, concurrent, convergent, and face validity. To have a better insight into what they refer to, the types of validity will be defined briefly below.

Content validity is defined as a non-statistical validity based on a systematic analysis of the test content to determine if it contains an adequate sample, namely, all major aspects covered in suitable proportions, of the target domain (Davies et al., 2002). In other words, if the content of a test includes a representative sample of the language skills, structures and so forth with which it is concerned, it is said to have content validity (Hughes, 2000, p. 22). Construct validity is another crucial part of assessment tools. According to Hughes (2000, p. 26), a 'construct' is "any underlying ability which is hypothesized in a theory of language ability". In a speaking test, such an ability may be, for instance, being able to ask for permission. Therefore, for a test to have construct validity, it should measure just the ability which it is supposed to measure. Concurrent validity, which is a subcategory of criterion-related validity (Hughes, 2000, p.23), is defined by Davies et al. (2002) as "the type of validity concerned with the relationship between what is measured by a test (usually a newly developed test) and another existing criterion measure". Thus, a

test is said to have concurrent validity if it correlates highly with another accepted measure. Convergent validity is related to the similarity between two or more tests which are claimed to measure the same underlying ability. This can be confirmed via a comparison of scores attained by a group of test takers on different tests (Davies et al., 2002). Finally, a test is said to have face validity if it looks as if it measures what it is meant to measure, as perceived by a person reviewing it (Davies et al., 2002; Hughes, 2000).

Reliability

Defined as "the actual level of agreement between the results of a test with itself or with another test" (p.168), reliability has three subcategories: *parallel forms, split half, and rational equivalence reliability estimates* calculated via selection of specific items*, test-retest reliability* checking whether a test would give consistent results when administered again in different conditions, and *inter-rater reliability* checking for the level of consensus between two or more independent raters (Davies et.al., 2002).

As Fulcher (2003) proposes, assessment of oral skills cannot yield entirely reliable scores, as the process is dependent on raters who will be influenced by numerous uncontrollable factors. Hence, test takers are likely to receive inconsistent scores due to the changing attributes of the raters. Brown (2004, p. 21-22) points out the distinction between intra-rater and inter-rater reliability and identifies more types of reliability influencing the overall reliability of a test: student reliability, which can be threatened by temporary illness, fatigue, or anxiety; test administration reliability, which can be threatened by external factors such as background noise; and test reliability, which depends on the inherent characteristics of a test such as being too

long. The literature on speaking tests has mostly focused on rater reliability. McNamara (2008, p.37) asserts that rating is necessarily subjective, that is, it is not only a reflection of the candidate's performance but also of the rater's characteristics, and adds that it always contains a significant degree of chance, no matter what is done to increase objectivity. Supporting this view, the findings of Lumley and Brown (as cited in McNamara, 2002) suggest that interlocutor behaviors can hinder or help candidate performance, and Lazaraton (1996) mentions a number of interlocutor behaviors that might affect the performance of the test takers in either direction. Among the precautions to be taken or points to be considered to retain reliability are: taking adequate samples of behavior, not permitting candidates excessive freedom, writing unambiguous items, giving clear and explicit instructions, making sure that tests are legible, presenting the questions in formats and with testing techniques candidates are familiar with, supplying a standardized and non-distracting environment for administration, using items that allow utmost objectivity in scoring, comparing candidates as directly as possible, giving a detailed scoring key, training the raters, determining acceptable responses and appropriate scores before scoring, scoring performances anonymously, and employing several independent scorings (Hughes 2000 p. 36-42). Finally, Fulcher (2003) notes that reliability is one of the major drivers of research into semi-direct tests of speaking, since semi-direct tests are seen as promising tools likely to yield more reliable results in the scoring of speaking tests. In support of this view, Galaczi (2010) argues that the role of interviewer variability in delivering the test, and the influence of rater variability in scoring the test, are reduced in computer based oral assessment.

In the present study, the face-to-face speaking assessment (FTFsa) and the computer-assisted speaking assessment (CASA) will also be examined in terms of their validity and reliability. Drawing on the growing evidence which favors direct over semi-direct in terms of validity, Cheng (2008) notes that it is unclear whether semi-direct tests can replace direct tests, yet there are studies supporting the view that semi-direct tests are reliable. The remainder of this section will review studies which have investigated the validity and reliability issues regarding semi-direct testing.

Being experienced in conducting the face-to-face OPI (Oral Proficiency Interview) and training the tape-based SOPI (Simulated Oral Proficiency Interview) raters, Kuo and Jiang (1997) compared these two forms of oral proficiency tests, examining them in terms of test administration, response elicitation, and rating procedures. The two tests, as examined by the authors, were found to be valid and reliable but to have different characteristics, and thus different advantages and disadvantages depending on the environment in which they are utilized. For instance, with better measured and controlled results, the SOPI was reported to be more reliable, though at the sacrifice of the human interaction element, whereas there is too much interviewer discretion in the OPI. The SOPI was said to be a more appropriate option where there are numerous interviewees but an inadequate number of raters, or where a uniform test is needed for a large group of test takers. On the other hand, the OPI was noted to be beneficial when human interaction, test adaptability, and personal information besides language ability were of concern. The authors therefore recommended choosing the appropriate test by considering the needs of the institution in which it would be used.

In an attempt to find whether direct and semi direct speaking tests can be scored reliably by different raters and whether they produce the same scores for any examinee, Stansfield and Kenyon (1988) administered two forms of a taped test and a live interview in Portuguese to 30 participants. Both test formats had questions regarding personal conversation, giving directions, detailed description, picture sequences, topical discourse, and different situations. The analyses showed that the inter-rater reliability was .95 for the taped speaking tests and .94 for the live interviews, which means that inter-rater reliability was not adversely affected by the semi-direct mode. The parallel-form reliability scores found conducting two different but parallel semi-direct tests ranged between .93 and 99, indicating that the tests drew uniformly challenging samples of speech, as the researchers suggested. Finally, the semi-direct test of speaking was claimed to be a valid test since the scores from it correlated highly with scores from the face-to-face live interview (.90 at least). Stansfield and Kenyon conducted a similar study in 1992 examining a semi-direct test of Indonesian speaking proficiency with similar results.

Qian (2007) compared two English proficiency tests - the English Test of the Graduating Students' Language Proficiency Assessment (GSLPA) and the Academic Version of the International English Language Testing System (IELTS) - in an attempt to examine the discriminating power of each test and to determine whether or not the speaking and writing components of the two different tests measure the same areas of language knowledge and skills. GSLPA's speaking component, in the form of a semi-direct test, takes place in multimedia language laboratories, whereas the speaking component of the IELTS is conducted in face-to-face format. The participants were a voluntary sample of 243 final-year students from 17 academic

departments at HKPU (Hong Kong Polytechnic University), who sat for both the GSLPA and the Academic Modules of the IELTS within a month. With regard to the speaking component, results indicate that GSLPA speaking scores distinguish candidates' abilities more clearly than the corresponding scores on the IELTS: although there are nine score bands in the IELTS, only bands 4-8 are used for scoring. The GSLPA scores are spread over a wider range and they are more evenly distributed. Nevertheless, IELTS overall scores, generated from writing, speaking, reading and listening sub-scores, have a discriminating power similar to that of GSLPA. The correlation between the scores on the GSLPA and IELTS speaking components is also fairly strong (0.69, $p<.01$, two tailed). The R^2 values indicate that 52% of the constructs of the two speaking subtests are distinct from each other and test different areas of knowledge, which is reasonable as the two tests have different purposes by nature.

Xiong, Chen, Liu, and Huang (as cited in Cheng, 2008) carried out a study in an attempt to find an alternative way of conducting a large-scale speaking test. The test takers were given a semi-direct oral test where they responded to prompts from a tape. Three different analytic rating scales (an ability scale, an item scale, and a holistic scale) were used to evaluate each student's performance to ensure the reliability of the test score. The scores from the three scales were reported to correlate highly. The researchers commented that the students demonstrated their actual oral language proficiency through the semi-direct test, counting on the fact that a high correlation was observed between students' ranking in class and the three scores. Therefore, the semi-direct test was considered by the researchers to be a reasonable alternative to direct tests.

In an investigation of the attitudinal reactions of test takers to different formats of oral proficiency assessments in Spanish, Arabic, and Chinese, Kenyon and Malabonga (2001) looked at the correlation of scores obtained in each test mode. A total of 55 students participated in the study. The students taking the Spanish tests took three types of tests: a tape-mediated Simulated Oral Proficiency Interview (SOPI), a Computerized Oral Proficiency Instrument (COPI), and the face-to-face American Council on the Teaching of Foreign Languages (ACTFL) Oral Proficiency Interview (OPI). The participants taking the Arabic or Chinese tests were administered the SOPI and the COPI only. The correlation between the students' scores in the SOPI and the COPI was .95, in the COPI and the OPI, it was .92, and the SOPI and the OPI scores correlated at a level of .94, which means that the examinees scored very similarly across the tests.

Jeong (2003) explored the relationship between 144 Korean college students' electronic literacy, assessed through the Electronic Literacy Questionnaire (ELQ), and the scores they obtained on a multi-media enhanced English oral proficiency interview, where the test takers were required to respond to the prompts given by a computer and record their voices. The participants took both a face-to-face and a multimedia enhanced oral proficiency interview utilizing d-VOCI (digital-Video Oral Communication Instrument). The researcher argued that the d-VOCI assessed not only linguistic knowledge but also communicative competence. Although both tests were supposed to share the same construct, a correlation of .30 showed that the relationship between the scores gained in the two modes was weak and low in a practical sense. Jeong suggested that this might have resulted from the low inter-rater reliability in the face-to-face test (.64) whereas d-VOCI established an inter-rater

reliability of .90. As for the results of the ELQ, a positive moderate relationship was found between the electronic literacy and the oral proficiency of the test takers.

Wigglesworth and O'Loughlin (1993) investigated the comparability of live interview and tape-based versions of a test with the participation of 83 candidates to find to what extent the test items were of similar difficulty, whether the test takers perform similarly on both modes, and to what extent their scores on each mode compare to the ratings obtained in a well established test. Performances on each mode were rated by two trained raters. The results revealed that both the live and tape based modes had a high degree of concurrent validity (.87 and .89) when compared to another well established test. Moreover, it was found that the candidates performed similarly on the two tests and that the test items were of similar difficulty.

O'Loughlin (2001) investigated the equivalence of direct and semi-direct tests in both theoretical and practical terms by examining the oral component of the *access:* test (the Australian Assessment of Communicative English Skills), administered around the world between 1993 and 1998. The researcher conducted the study in the form of an instrumental case study with the aim of examining the construct validity of the two alternative modes of speaking tests. O'Loughlin's purpose was to determine whether they, in fact, measured the same kind of ability and whether this ability was measured with equal precision in each mode. He concludes, via a multifaceted validation process, that the spoken interaction of two or more people is mutually constructed and therefore basically different in its character from communication with a machine. O'Loughlin therefore cautions against using the direct and semi-direct forms of the test interchangeably because even small

reactive tokens such as *hmm*, *yes*, *right* made a measurable difference to the character of the test taker's response.

Xi (2008) conducted a study to provide criterion-related validity evidence for ITA (international teaching assistant) screening decisions based on TOEFL IBT Speaking scores and to evaluate the adequacy of using the scores for TA assignments. The researcher investigated the relationships between scores on a TOEFL Speaking test and scores on criterion measures, namely, locally developed teaching simulation tests used to select ITAs. The participants were 253 ITAs from four different universities which were selected as they had established procedures to select ITAs: University of California, Los Angeles (UCLA); University of North Carolina, Charlotte (UNCC); Drexel University (Drexel); and University of Florida at Gainesville (UF).The tests used at these universities are performance based tests that attempt to simulate language use in real instructional settings. Some of the participants received one of the two forms of the TOEFL IBT Speaking test containing six speaking tasks at the beginning and then took the local test at their university, while others took the local test first and the other form of the TOEFL IBT Speaking test later. The use of the TOEFL Speaking test for ITA screening is supported by the findings as TOEFL Speaking scores were reasonably correlated with most scores on the local ITA-screening measures. According to the observed and disattenuated correlations respectively, the TOEFL Speaking scores had the strongest relationship with the speaking test scores at UCLA (.78/.84) and the non-content-based test at UNCC (.78/.93), weaker relationships with the speaking test scores at Drexel (.70/NA) and the content-based test at UNCC (.53/.58), and the weakest relationship with the UF Teach Evaluation scores (.44/.72).

The diversity in the results of the studies on validity and reliability of the semi-direct speaking tests reported above might have interacted with numerous factors, one of which may be test takers' perceptions and the relationship between their perceptions and their test scores. The next section will provide detailed information about the studies conducted with the aim of shedding light on to test takers' perceptions of the speaking tests administered in semi-direct mode.

Perceptions of Test Takers

Among various factors that might affect individuals' test performance, their perceptions of the tests have been of interest to the researchers.

Investigating the development and validity of the Portuguese Simulated Oral Proficiency Interview (SOPI), the semi-direct tape-based version of the Oral Proficiency Interview (OPI), Stansfield and Kenyon (1988) found that the tape-based semi-direct format was less popular among the test takers than the face-to-face OPI. A total of 30 subjects were asked to complete questionnaires addressing their perceptions of the two test types. Although they achieved approximately the same scores in both types of tests, the majority was reported to perceive the live format as less difficult. Looking at the comments by the participants, Stansfield and Kenyon concluded that this was probably a reflection of the face-to-face testing mode, which seemed more natural, rather than a reflection of the technical quality of the taped test. Indeed, the participants were positive about the content, technical quality and the ability of the semi-direct test to predict their oral language proficiency. Nevertheless, the fact that the mode of testing was unfamiliar and speaking into a tape seemed 'unnatural' to the participants resulted in a greater perceived difficulty and more nervousness than the face-to-face format. A similar study (Stansfield et al., 1990)

found that 73% of the participants felt that their maximum level of Portuguese speaking ability had been probed by both the live interview and the SOPI, yet 90% perceived the taped test as more difficult, 70% felt more nervous in the SOPI, and 86% preferred the live interview. In short, the majority of the participants favored the face-to-face speaking test over its semi-direct counterpart.

Kenyon and Malabonga (2001) explored examinee attitudinal reactions to taking different formats of oral proficiency assessments across three languages: Spanish, Arabic, and Chinese, as mentioned above. Of 55 students in the study, 24 participating in the Spanish study took three types of tests: the SOPI, the COPI, and the OPI. The 15 participants taking the Arabic tests and 16 taking the Chinese tests were administered the SOPI and the COPI only. The examinees scored very similarly in all tests. Following each test, the examinees completed a questionnaire on their attitudes towards and perceptions of that test, and finally they were administered another questionnaire asking them to compare the test modes. Comparisons were made in six categories: opportunity to demonstrate strengths and weaknesses in speaking, test difficulty, test fairness, nervousness, clarity of instructions, and representativeness of the performance. Kenyon and Malabonga report that both the SOPI and the COPI were perceived as equally fair and clear while the participants found SOPI more difficult. On the other hand, when the perceptions regarding the semi-direct format were compared to those related to the OPI, it was found that there was still a definite preference for the OPI as the participants stated that it gave a better opportunity to demonstrate their speaking ability. The researchers concluded that neither of the technologically-mediated tests could replicate the interactive,

conversational and personal nature of the face-to-face interview for the Spanish examinees.

Investigating 144 Korean college students' attitudes towards the d-VOCI (digital-Video Oral Communication Instrument), Jeong (2003) administered one face-to-face and one semi-direct test of speaking proficiency, namely, d-VOCI in an EFL setting. As reported before, the correlations between the scores from the two tests were weak. This inequality was also observed in the students' attitudes towards the test mode as revealed by their responses to a 30 item subscale in the Electronic Literacy Questionnaire. The d-VOCI was reported by 83% of the students to be promising in that it would improve their English proficiency and 90% stated that they liked that their performances were scored by qualified OPI raters. Nevertheless, 70% of the participants noted they would prefer the live format as it was more authentic and interactive.

Qian (2009) compared the popularity of two testing modes, namely, direct and semi-direct modes, by analyzing the reactions and perceptions of a group of test takers who had just sat for a direct test as well as a semi-direct test in a university setting in Hong Kong. The direct test consisted of the speaking component of the International English Language Testing System (IELTS) and lasted 10-14 minutes, whereas semi-direct testing was represented by the speaking component of the Graduating Students' Language Proficiency Assessment–English (GSLPA) and lasted 40 minutes. A total of 243 final-year students from 17 academic departments volunteered to take the tests. Following the tests, 186 of them responded to a questionnaire asking them to report on their reactions to and perceptions of the two tests. The survey also included open-ended questions added to the questionnaire as

well as some follow-up interviews to receive test takers' comments on test content relevance, test design, test mode, test usefulness, and test takers' preferences. Qian drew on Krashen's affective filter theory while interpreting subjects' perceptions regarding the tests and commented that if a test taker's state of mind or disposition was affected by the testing mode in some negative way, the affective filter might also be up to interfere with his or her test performance. The results indicated that the number of respondents who strongly favored IELTS and found it more authentic (61, or 33%) exceeded the number of those in favor of the GSLPA (18, or 10%). On the other hand, 41% of the respondents were positive toward both testing formats and 58% actually did not show a particular preference. The researcher interpreted this as a signal for the promising future of semi-direct tests.

In a study examining test takers' attitudes towards the TOEFL IBT in China, Colombia, Egypt, and Germany, Stricker and Attali (2010) collected data through TOEFL IBT and a questionnaire on attitudes completed by 762 of the test takers. It was found that the mean scores for TOEFL acceptance as measured by a subscale in the questionnaire was moderately positive in China, Colombia, and Egypt while they were neutral for Germany, which means that the participants in Germany favored the TOEFL IBT less, in general. Moreover, fewer favorable responses came from Germany and Egypt regarding the reading section and fewer people favored the speaking section in Germany. The speaking section was the least admired one in all countries. The researchers pointed out that the reason the speaking section was favored less might be its absence of interaction. On the other hand, most test takers still had positive feelings towards the TOEFL IBT in general. Test performance and attitudes were reported to be weakly related, and no significant relationship was

found between the attitudes and speaking performance in particular. TOEFL acceptance was found to correlate with computer attitudes and familiarity in some countries, which might indicate that test mode has an effect on attitudes.

Anxiety and Performance in Oral Assessment

Anxiety Types and Definitions

One of the research questions of the present study focuses on the relationship between different types of anxiety and performance of the test takers in particular speaking tests. Therefore, giving a definition and a brief description of types of anxiety followed by studies investigating the abovementioned relationship would shed light into the issue.

Birjandi and Alemi (2010) point out two classifications of anxiety. The first comprises trait, state and situation-specific anxiety, whereas the second involves facilitating and debilitating anxiety. Trait anxiety is said to be a stable aspect of one's personality, state anxiety is the nervousness felt at a specific moment in a particular setting, and situation-specific anxiety pertains to the negative feelings experienced in a specific type of situation such as speaking in public or taking an examination. As for the second type of distinction, facilitating anxiety is the type which "mobilizes resources to accomplish a task" whereas debilitating anxiety is an excessive amount of anxiety which hinders learning (Birjandi et al., 2010; Ehrman, 1995). Foreign language anxiety, defined by Horwitz, Horwitz and Cope as "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process" (1986, p.128), falls under situation-specific anxiety as Birjandi et al. suggest, and it can either be

debilitating or facilitating. Communication apprehension, fear of negative evaluation, and test anxiety are the subcategories constituting foreign language anxiety (Horwitz, Horwitz & Cope, 1986).

The anxiety levels of test takers may differ due to various factors, i.e. in a computer assisted speaking assessment, they may feel more nervous about the test because everything they say is recorded and no gestures or expressions can be used, which leaves speaking as the only channel, or they may be discouraged when they encounter a live interviewer in a face-to-face speaking test. To find how tests takers' anxiety levels are affected in different situations, a number of researchers have studied the relationship between speaking tests and the anxiety levels of the examinees.

The Relationship Between Speaking Test and Speaking Anxiety Level, and Performance in Oral Assessment

One of the possible factors that may pose a threat to the demonstration of a test takers' full competency is anxiety. The literature reviewed suggests contradictory findings as to the effects of anxiety on test performance.

Guo (as cited in Cheng, 2008) tested ten final year English majors in three situations: (a) recording their opinions of a topic on a tape (b) talking to some freshmen in a casual environment; and (c) talking to a tester in an office. The purpose was to explore the correlation between their motivation and oral performance in each situation, so the students also completed a questionnaire on motivation. Examining the length of natural pauses and frequency of unnatural pauses, the researcher found that there was a high correlation between motivation

and length and frequency of pauses in the first and third situations. Guo suggested that the pressure felt by students in different situations led to various degrees of anxiety and this affected their fluency. The participants were most fluent in the second situation, which was a casual environment, which, in turn, means that the tape-based version (first situation) caused more anxiety. Although the limited number of participants hinders the generalizability of the study, the researchers' suggestion of considering test takers' affective factors while developing oral tests is one that deserves attention.

In a study comparing a tape-mediated test of speaking which had four tasks and a face-to-face test composed of a warm up and two tasks with the participation of 37 candidates in Finland, Luoma (1997) found that the participants felt more anxious in the tape-based form, yet there was only a moderate amount of anxiety in the face-to-face mode. Although the participants were not entirely unhappy with the taped version as 85% of them thought that the test corresponded to real life situations, they complained about the lack of interaction, and hearing others' voices while trying to respond. Together with these factors, the researcher argued that the absence of experience with taped tests could have caused more anxiety in the taped semi-direct speaking test. On the other hand, the test takers were glad to have someone listening to them in the face-to-face speaking test because it raised a feeling of authenticity and success. It is important to note, though, that the test takers produced more linguistic but less content-oriented responses in the taped version, which suggests that the linguistic accuracy surprisingly increased as the anxiety increased.

Phillips (1992) reported on a study that examined the influence of students' anxiety on performance on a French speaking test and explored the attitudes of

highly anxious students towards that exam. A total of 44 students studying intermediate French at Southwestern University individually took a face-to-face French speaking test. They also completed the Foreign Language Classroom Anxiety Scale (FLCAS) developed by Horwitz et al. (1986), and six highly anxious students were interviewed for an insight into their attitudes towards the test. A moderate negative correlation (-.41) showed that there was a negative relationship between the performance on the oral examinations and anxiety level. The results also revealed that the highly anxious participants were inclined to say less, to produce shorter communication units, and to use fewer dependent clauses and target structures than low anxiety students. As for their attitudes, having experienced high levels of anxiety, both high and low proficiency students interviewed found the assessment very unpleasant and reported getting frustrated at forgetting what they actually knew, going blank, and feeling panicky. Briefly, this study supports the view that anxiety has a negative effect on speaking performance.

Phillips (2005) investigated the impact anxiety might have on Chinese students' speaking proficiency. A total of 62 students taking an English course at HKU were given the FLCAS first, and the top and bottom 25 percent of the students were classified as having high or low anxiety; four from each group were later chosen for further participation. As a part of the course, all of the students gave a short presentation which was video-taped and performances from which were rated by three instructors. Finally, the participants were asked to watch their recorded presentation, comment on it, and answer some interview questions. Phillips found no significant correlations between the scores they received and their anxiety levels as probed by the FLCAS and the interviews. The researcher interpreted the findings as

unclear as students with similar levels of anxiety got both high and low grades, meaning that showing that high anxiety does not essentially mean low oral proficiency.

Oya and Greenwood (2004) investigated the relationship between anxiety levels and personality with regard to extroversion and neuroticism of 73 intermediate Japanese tests takers and their English speaking proficiency. The data were collected through the use of the Maudsley Personality Inventory (MPI), the Japanese version of the Spielberger State and Trait Anxiety Inventory (STAI), and a story-retelling task where students were asked to order 6 pictures and tell a relevant story about them into a microphone. The recorded performances of the test takers were then scored by three raters independently in terms of accuracy, fluency, complexity, and global impression, with high levels of inter-rater reliability. The results indicated that only the global impression of the participants' oral performance significantly correlated with their extraversion scores, but no significant correlation was found with neuroticism scores. It was also found that that accuracy as measured by accurate clause rate was significantly negatively correlated with the participants' anxiety scores (-.23), which means that as the anxiety levels went up, the accuracy of their speech decreased.

Obviously, anxiety has some effect on speaking performance of the test takers. Neverthelesss, it is not intelligible to explain all differences in oral performances in semi-direct speaking tests with merely one variable. The next section will focus on another possible factor, test takers' computer attitudes, that might have an effect on oral performance in such tests.

The Relationship Between Computer Attitutes and Performance in Oral Assessment

Chapelle and Douglas (2006) draw attention to the possible effect of computer familiarity on test takers' performance on computerized tests. They assert that "the lack of familiarity might result in shaken confidence that could negatively affect the performance on the task" in spite of a strong knowledge of the construct tested in a test situation (p.43). They add that it could more critically result in noncompletion of a set of items due to the loss of time resulting from numerous attempts to figure out how to respond to the items in a computerized environment. Therefore, the researchers recommend that those who investigate all types of language tests should pay attention to the fact that the computer mode of delivery may influence performance. With a similar perspective, a few studies have investigated the effects of computer familiarity on test takers' performance on speaking tests, in particular.

Though not related to speaking, a study by Kirsch, Jamieson, Taylor and Eignor (1998) explored the relationship between computer attitudes and proficiency scores obtained in a paper based TOEFL as an answer to the concern that a computer based test might confound English proficiency with computer familiarity. A total of 89,620 candidates taking the TOEFL test completed a 23 item questionnaire, the Computer Familiarity Questionnaire developed by the researchers, which involved items related to computer access, attitudes, experience, and related technology. The participants were classified into three groups; high, moderate, or low familiarity, via the questionnaire and the questionnaire, from each participant was matched to their TOEFL test scores looking for a relationship between them. The results revealed that

the average difference in test scores, between the low and high computer familiarity groups was around 25 points; more specifically, the candidates with high computer familiarity were expected to get an average TOEFL score of about 536, while it was around 510 for the low-computer-familiar candidates, which is a considerable difference.

In a study researching whether the relationship between test anxiety and test performance was the same in the paper-based or computer-adaptive Graduate Record Exam (GRE), Powers (1999) compared the scores from the test with the results from a computer attitudes inventory and a test anxiety inventory. The researcher found the test anxiety to be very similar in the two testing modes. It was also found that neither computer anxiety nor computer confidence, the sub-components of the computer attitudes inventory, interacted with test mode. Nor was a significant relationship found between computer attitudes and the scores from the computer-adaptive GRE.

Goldberg and Pedulla (2002) also investigated the relationships between test mode and computer familiarity with test performance on the GRE. In contrast to Powers (1999), the researchers found differences between the performances in the two testing modes. The examinees in the paper-and–pencil group performed better than those in the computerized test group. The level of computer familiarity was found to relate to the Quantitative and the Analytical subtests of the GRE, where the higher computer familiarity group outperformed the lower computer familiarity group.

On the other hand, utilizing a 100 item multiple choice test related to the content in a business class and learner self report information about 105 freshman

undergraduates' characteristics, Clariana and Wallace (2002) found that the computer-based test group outperformed the group taking the paper based test. The researchers further report finding that computer attitudes were not related to differences in performance.

In short, the studies related to the influence of computer familiarity or computer attitudes on test performance provide contradictory results. It is also important to note that none of the studies reported above investigated the relationship between computer attitudes and oral performance in a computerized test, which is the focus of the present study.

Conclusion

In this chapter the importance and qualities of the speaking skill and speaking assessment, the history, characteristics, validity and reliability of semi-direct oral assessment, and studies related to test takers' perceptions and anxiety levels as well as their atttitudes depending on the speaking test mode were reported in the light of the relevant literature.

As seen in the discussion of the development of semi-direct oral assessment, numerous attempts have been done to create a speaking test to conveniently assess the speaking ability of large number of test takers; however, the tests developed have aimed at assessing general oral proficiency mostly using tape-recorders, unlike the persent study which aims at utilizing a computer-assisted speaking test as an achievement test using a simple technology which can be installed and used with ease. Moreover, unlike the studies reported here, this study comprises two detailed questionnaires prepared with the purpose of getting a deeper insight into the test

takers perceptions of the test modes, which are also intended to find out about the test takers' test-mode-related anxiety levels. It was seen that the literature mainly focused on the validity and reliability of the semi-direct speaking tests in addition to few studies which sought for the attitudes of test takers towards semi-direct tests. Nevertheless, the studies searched neither for the relationship between test takers' speaking anxiety, speaking test anxiety levels, nor their computer attitudes and their performance on the semi-direct speaking tests, which might indeed be important aspects influencing oral performance. Finally, this study aims at finding whether there is a difference between the speaking test scores, perceptions, attitudes, or anxiety levels of test takers and their proficiency level by looking at participants from two proficiency levels, pre-intermediate and intermediate.

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

CHAPTER III: METHODOLOGY

Introduction

This exploratory study focused on a comparison of face-to-face (FTFsa) versus computer-assisted speaking assessment (CASA) in an EFL setting in an attempt to shed light on the relationship between the mode of speaking assessment and student performance, student perceptions of the modes, the anxiety levels, and the computer attitudes of the students. The study aimed to address the following research questions:

1. How is the speaking performance of the pre-intermediate and intermediate level test takers at tertiary school affected by the test mode being either the face-to-face (FTFsa) or the computer-assisted speaking assessment (CASA)?
2. What are the test takers' perceptions of oral assessment?
    a. What are their perceptions of the FTFsa?
    b. What are their perceptions of the CASA?
3. What is the relationship between the anxiety levels of the test takers and the test mode?
    a. What is the relationship between speaking/speaking test anxiety, and test scores?
    b. What is the relationship between speaking/speaking test anxiety, and students' perceptions of FTFsa and CASA?
4. What is the relationship between the computer attitudes of the test takers and the test mode?

   a. What is the relationship between students' computer attitudes and test scores?
   b. What is the relationship between students' computer attitudes and their perceptions of FTFsa or CASA?
5. Depending on the test mode, do the speaking performances, test-mode-related perceptions, and anxiety levels of the test takers at different proficiency levels differ?

This chapter describes the methodology of the study. The following subsections review the setting, participants, instruments, data collection procedure, and data analysis.

Setting

The research was conducted at Uludağ University School of Foreign Languages, in Bursa, Turkey. As for the choice of the institution, eligibility and needs were of primary concern. The school is in charge of giving compulsory or optional extensive English language education for students who have passed the university exam before they start their bachelor's education in their departments. The program lasts for one year and consists of three proficiency levels: elementary, pre-intermediate and intermediate. Students are put into groups based on the scores they got on a proficiency/placement test given at the beginning of the year. During both semesters, students are required to take achievement tests and at the end of the year they are asked to take an exit exam to demonstrate that they have completed the program successfully. To date, tests have been used to evaluate students' competency in grammar, vocabulary, reading comprehension, listening and writing, which is in

line with the courses offered. The 2010-2011 academic year is the first time speaking courses have been integrated into the curriculum. Therefore, no tests were allocated to evaluate oral proficiency prior to the present study.

Participants

Four instructors, and a total of 75 students - four groups from two proficiency levels, pre-intermediate and intermediate - participated in the study in the beginning. Of all participants, nine students were discarded because they dropped from the language program, took only one of the tests, or answered only some of the questionnaires. Also, two students were left out from score comparison analysis because they took a grade of "0" in one of the tests since they did not answer the questions properly. However, their answers for the questionnaires were included in the study as they were answered independently from the tests. As a result, there were two groups of intermediate students with 19 students in each, and two groups of pre-intermediate students with 13 in one group and 13+2 in the other group (see Table 1). In total, the number of student participants was 66.

Table 1

Distribution of the Participants According to Levels and Groups

Level	Group	Number of Students	Instructor	Raters
Pre-intermediate	I	13+2	A	A-C
Pre-intermediate	II	13	A	A-C
Intermediate	I	19	B	B-D
Intermediate	II	19	B	B-D
Total		66	2	

The sampling was done on availability basis: At each level, attention was paid to choosing classes who shared the same instructor to achieve reliability, diminishing the chances of any differences in instruction as it might have affected the test results. For the rating process, the students' own class instructor and an instructor who does not give any courses to the participants functioned as raters (Table 1). The instructors were non-native speakers of English. Both the instructors and students were asked for their informed consent. All students were administered both the face-to-face (FTFsa) and Computer-assisted speaking assessment (CASA) and the questionnaires.

Elementary level students were not included in the study as their speaking classes focused merely on short-responses, making it impractical for that moment to require them to answer different types of speaking questions.

Instruments

Two types of data collection instruments, speaking tests and questionnaires, were used in the study. Some of the instruments were adapted from relevant literature while the others were created by the researcher. Each of the instruments is described in a separate section below.

Speaking tests

Eight speaking tests -four FTFsas and four CASAs- were developed in order to evaluate students' progress with regard to oral competency. The course contents at both levels were taken into consideration while preparing the questions. Both the FTFsas and CASAs were composed of a diverse range of questions in order to ensure that they were as communicative and real-life like as possible besides having face

and content validity. The researcher was inspired by the questions in previous studies (Clark, 1986; Stansfield, 1988, 1990, 1992) while preparing some of the questions, whereas some were created originally by the researcher based on the course content and level of the students. The tests had parts devoted to introducing oneself, commenting on a given situation, picture/graph description, topical discourse, situational discourse, simulated conversation, detailed description and discussion questions. To be precise, the questions mainly assessed specific aspects of spoken language introduced in the speaking classes which included giving a short presentation about a familiar topic, asking for advice, making recommendations, talking about one's personality, job and company profile, making requests and justifying them, talking about possibilities in a given situation, explaining a familiar concept, discussing an idea. In the FTFsa, the questions were read aloud or acted out by the interlocutors who followed the instructions on the paper that also included the questions. In the CASA, the instructions on how to respond to the questions and the questions were presented mainly in written format and the visual aids or listening materials were conveyed through the video screen embedded in each page. The responses were also recorded by clicking on the buttons on this video screen. (See Figures 1, 2, 3 & 4). For a sample of questions used in the tests, see Appendices A and B.

The FTFsa started with a warm-up question asking about personal information and continued with questions which were presented in a thematic order so as to ensure coherence and enhance authenticity. The instructors conducting the face-to-face speaking tests were both non-native speakers of English one whom was only the rater whereas the other was both a rater and the class teacher of the

examinees. It was both instructors' responsibility to conduct the interview, which means that the students were tested by someone to whose voice and style they were used to as well as someone unfamiliar in order to increase variability and thus, validity and reliability of the tests.

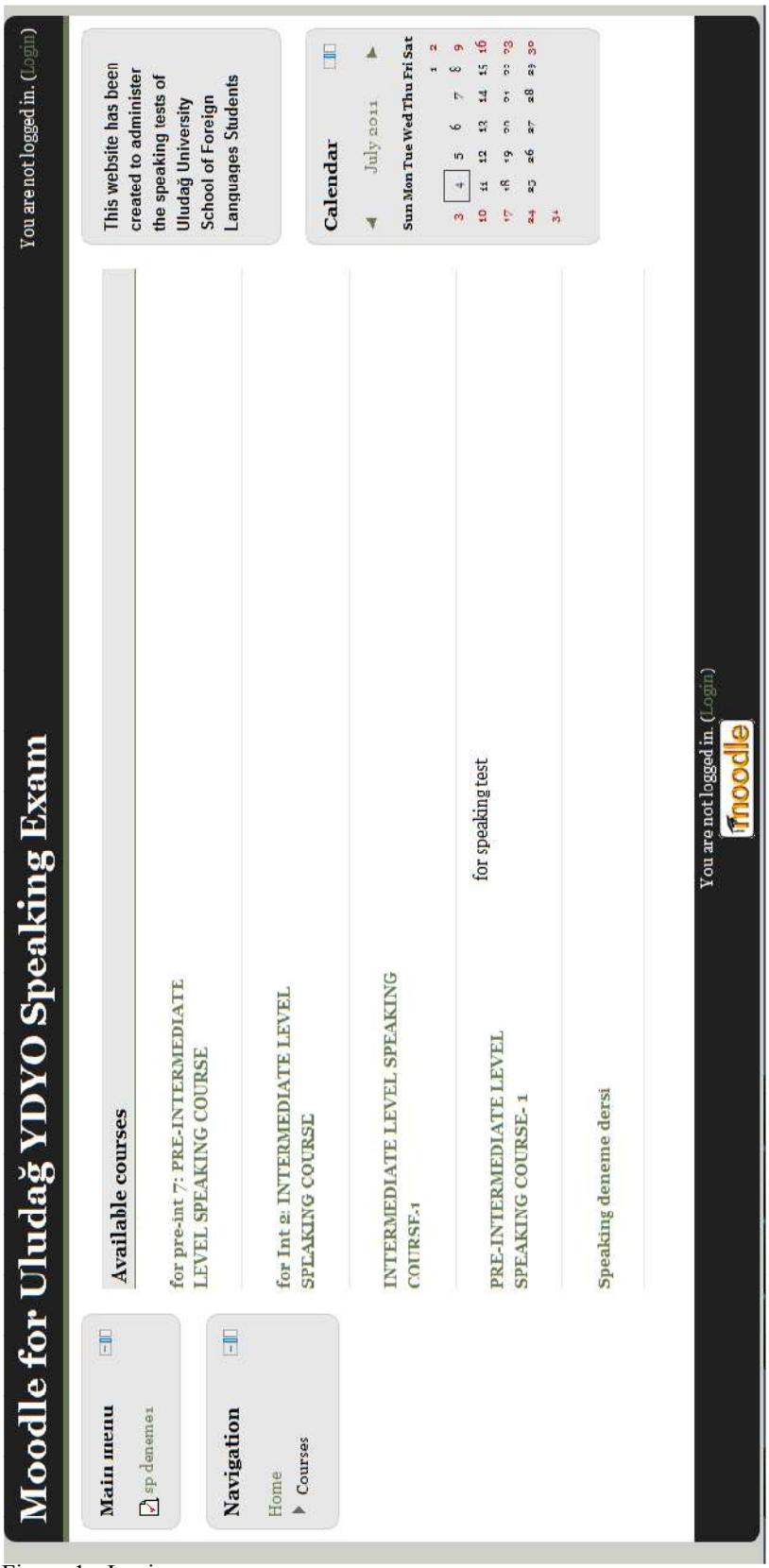

Figure 1 - Login page

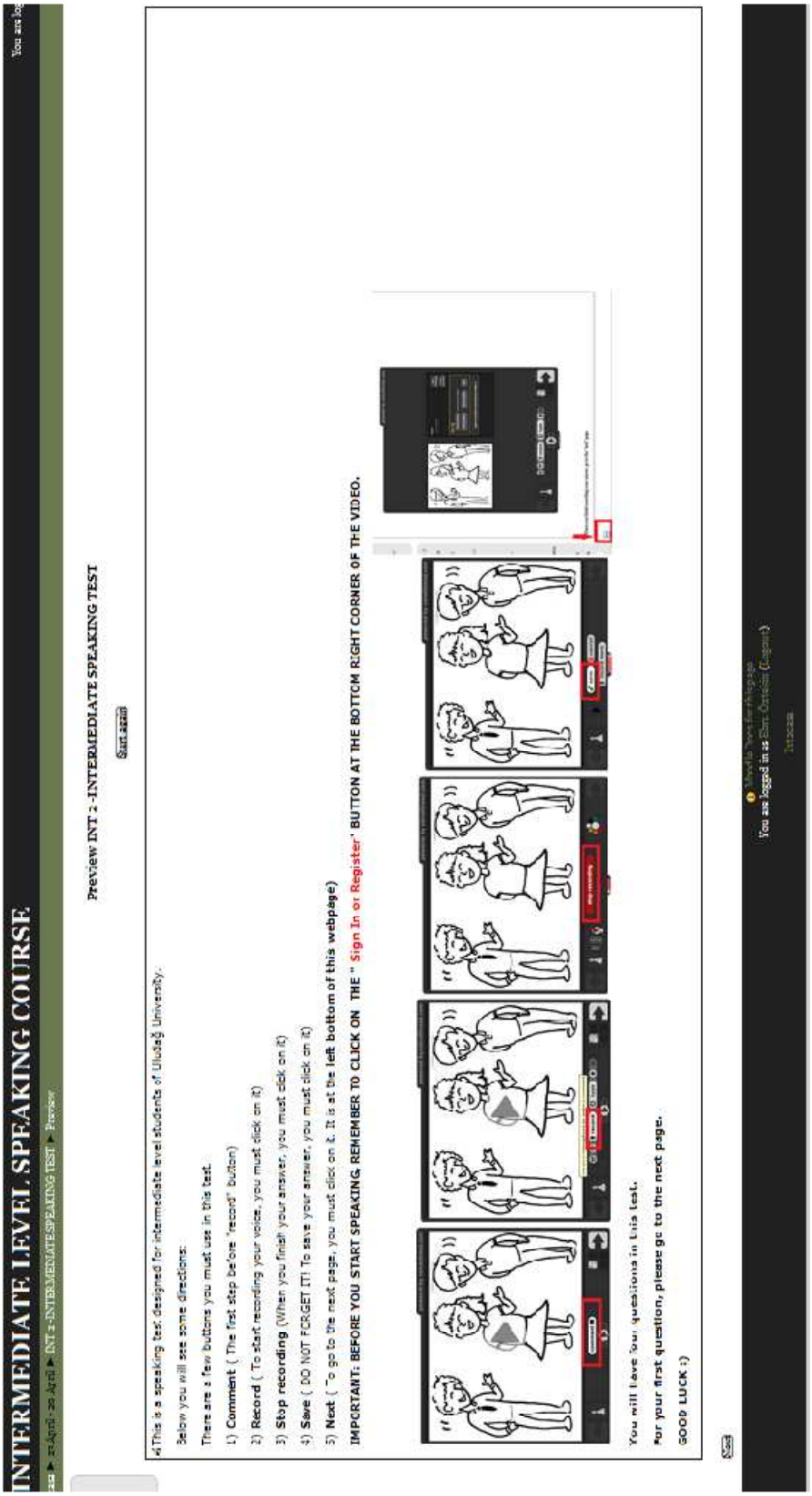

Figure 2 - First page of the CASA

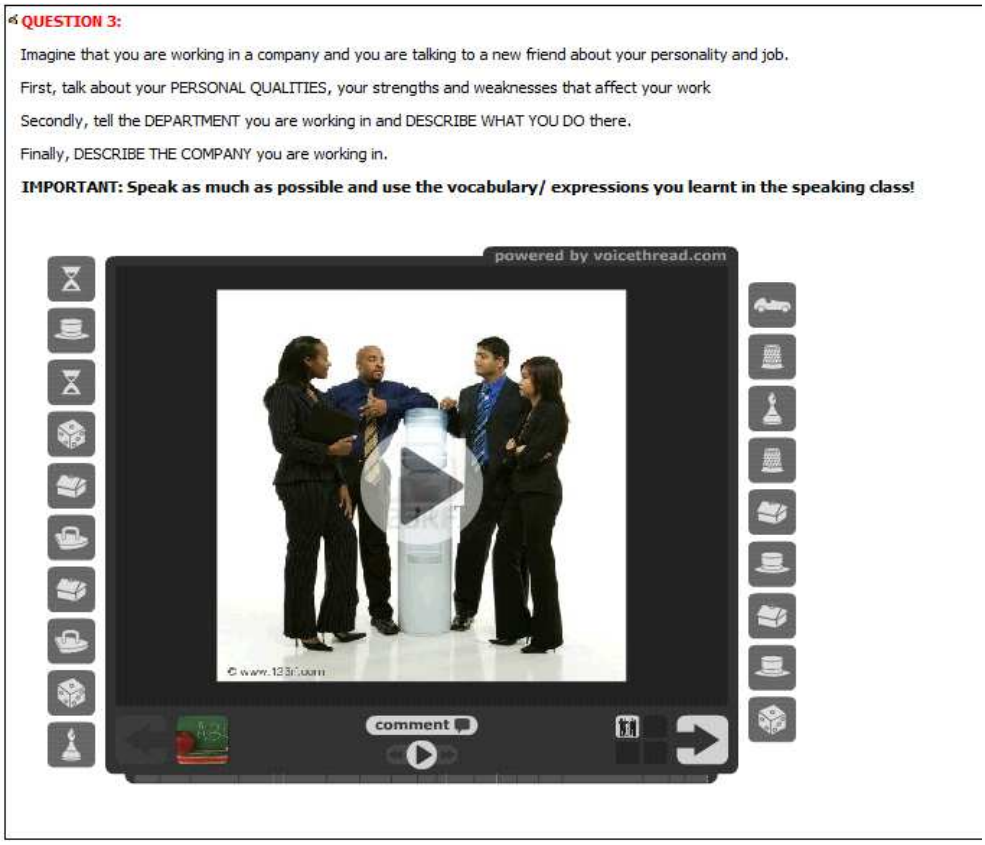

Figure 3 - Sample question in CASA

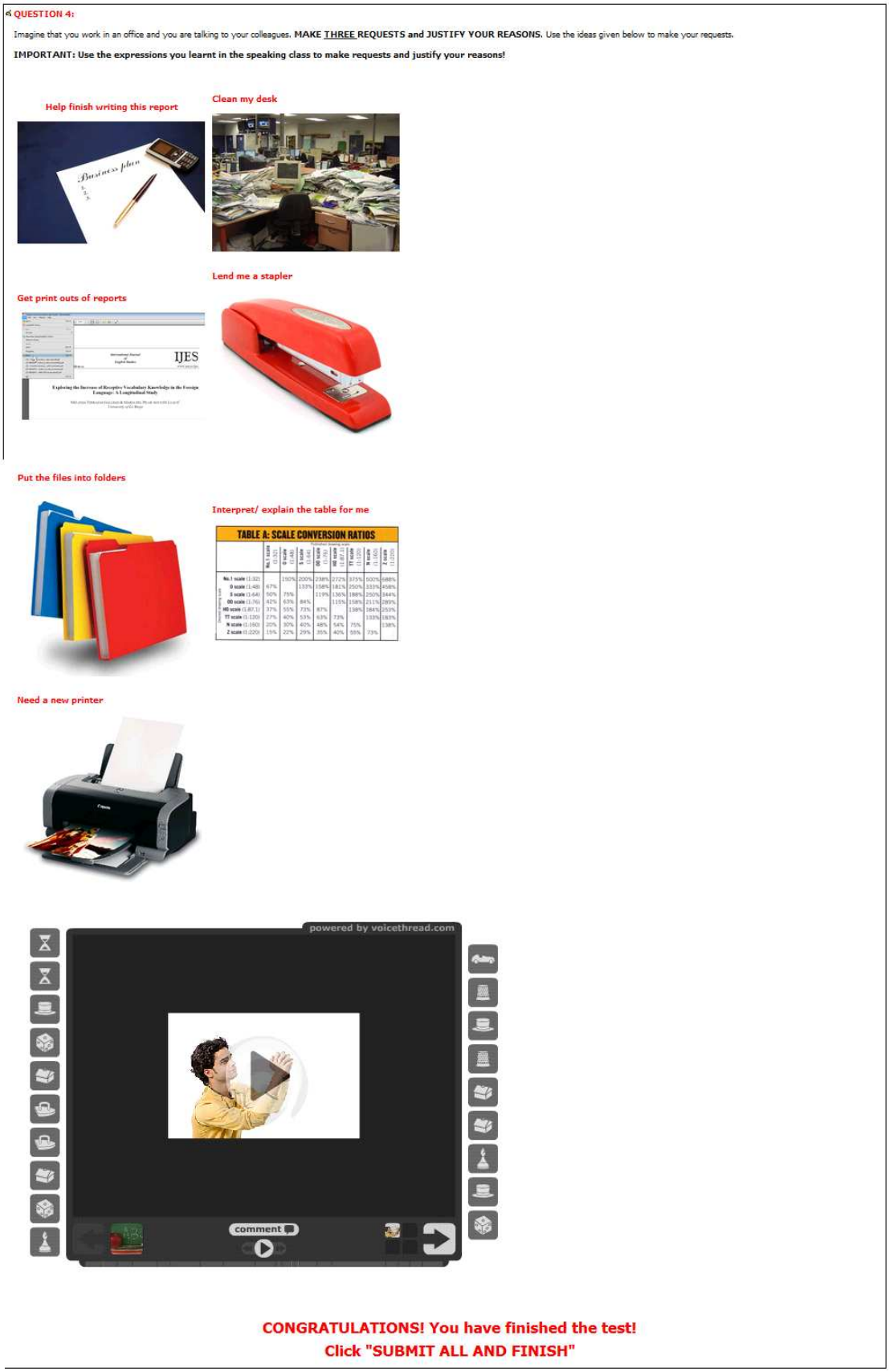

Figure 4 - Sample question 2 in CASA

As for the software used to create the CASA, a free learning management system (LMS), Moodle 1.9 from http://moodle.org/, and a commercial website, http://voicethread.com/ were chosen due to their availability, feasibility and convenience. Both software products were used to present test questions, record student responses for the CASAs, and store them. First, Moodle was installed in the researcher's computer with the assistance of an instructor experienced in setting up and utilizing the program. Second, necessary adjustments were made by the researcher so that she could create the tests and enroll the students to Moodle. The students were enrolled to VoiceThread and Moodle by the researcher and a list of usernames was prepared to be distributed to the participants later on. To create the questions, recorded versions of the questions that were also asked in the face-to-face test were uploaded to VoiceThread. Next, the VoiceThread links with the recordings were embedded under the relevant titles in Moodle. This was repeated for each of the four CASAs as well as the trial speaking test that was used to introduce the CASA to the participants. The Moodle pages consisted of instructions written in L2, and a media player screen where the test takers saw relevant pictures, listened to the recordings, and recorded their answers using the buttons on it. The computers in the computer laboratory the CASA would take place in needed some adjustments. Therefore, some of the computers were changed with new ones; problems with internet connectivity and other software were eliminated. New headphones and microphones were added to each computer. Moreover, it was ensured that each computer had the Adobe Flash Player ® 10, the compatible version, by updating an earlier version.

There was a one-month interval between the first and second speaking tests each participant took. The tests were conducted in a counter-balanced order to diminish the order effect. In other words, half of the test takers at each level were given the CASA first while the other half were given the FTFsa initially with the aim of avoiding results that would support the first or the second conducted test due to positive or negative practice effect. Table 2 shows the design of the study.

Table 2

Counter Balanced Design of the Study

	Pre- intermediate		**Intermediate**	
	Group I	**Group II**	**Group I**	**Group II**
Test I	FTFsa	CASA	CASA	FTFsa
Test II	CASA	FTFsa	FTFsa	CASA

Note. Group = the order the participants took the test

The scores on the speaking tests were given according to a five-section rating scale consisting of accuracy, fluency, pronunciation, vocabulary, and coherence/discourse sections accompanied by information such as date, name of the rater and the student, class, and test mode; and a document with band descriptors, - adapted from the the section titled "speaking criteria" in the website of the Language Leader text book published by Pearson Longman- providing detailed descriptions of the abovementioned sections and the corresponding scores out of five for each section (see Appendices C and D). The raters were also provided with a list of expected answers to increase the objectivity and reliability of the scoring (see Appendix E). The raters were required to check whether the test takers' responses matched the "expected answers" and give appropriate scores using the band descriptors. For instance, if the "expected answers" list had required the use of answers such as *It's a good place to go if…, It's handy for…, It's popular for…, You*

should definitely…, The place is well worth… etc. and the test taker had failed to use the required structures, she or he would have received a score of 0 for accuracy and vocabulary based on the requirements indicated in the band descriptors about accuracy and vocabulary.

Questionnaires

Three questionnaires, one of which had two versions, were used to gather data: a test perceptions questionnaire- one version for the CASA and one for the FTFsa, a Speaking Test and Speaking Anxiety Questionnaire, and a Computer Attitudes Questionnaire. All of the questionnaires had a cover page including a section informing the participants about the aim of the questionnaire accompanied by an informed consent form, a section where demographic information was sought, and an explanation of the Likert-scale included in the questionnaire. The questionnaires consisted of statements where the participants were expected to give scores on a five or six point Likert scale as well as open-ended questions at the end. Due to the fact that the participants were non-native speakers of English and might have had difficulty in apprehending the questionnaire in English, the Turkish versions of the questionnaires were administered. Figure 5 shows the sequence the questionnaires are administered in two groups taking different tests at the same time.

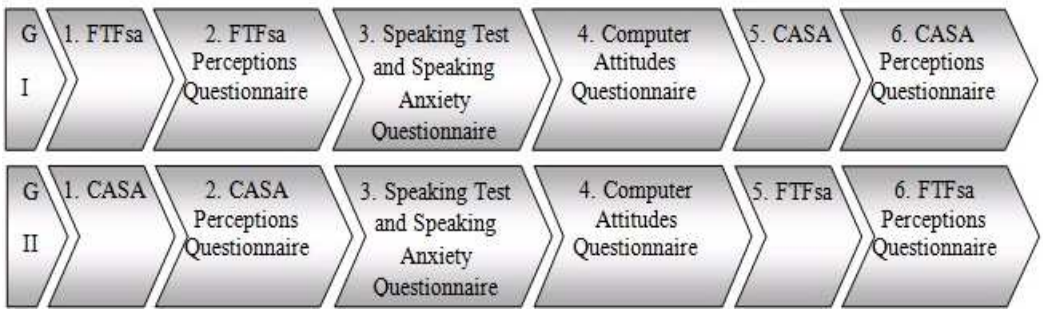

Figure 5 - Administration sequence of the questionnaires

Note. G = Group

CASA and FTFsa Perceptions Questionnaires

In order to collect information about the participants' perceptions of the two test modes, two separate questionnaires were prepared for each test mode (see Appendices F, G, H, and I). The questionnaires were partly adapted from Kenyon and Malabonga (2001) but most of the questions were originally generated by the researcher. Both questionnaires comprised 29 questions which were answered on a six-level Likert scale. The responses ranged from strongly agree =6, to strongly disagree=1. Each questionnaire also contained three open-ended questions (see Appendix J for responses) which also sought information about the test takers' perceptions of the test modes. While most of the questions were individual items, nine items comprised a test-mode-related anxiety subscale.

Initially, the wording of the first drafts of the questionnaires were checked by a group of students similar to the participants of the main study and some items were modified to clarify their meanings. Next, each group of test takers received the relevant questionnaire immediately after they took the CASA or the FTFsa so that the researcher could collect information on their real opinions and feelings about the test mode in question. The questionnaires were administered in Turkish in order to avoid any misinterpretations by the students. Students who did not submit the questionnaires were discarded from the study.

Speaking Test and Speaking Anxiety Questionnaire

The questionnaire comprised two parts focusing on two different types of anxiety: Speaking anxiety and speaking test anxiety. To explore the levels of these two types of anxiety, two well-grounded questionnaires were adapted and combined:

the Test Influence Inventory (TII) by Fujii (1993) and the Foreign Language Classroom Anxiety Scale (FLCAS) by Horwitz (1986). The speaking test anxiety subscale was composed of 23 items and the speaking anxiety questionnaire had 25 items, (which were followed by an open-ended question.) responses to which ranged from strongly agree=5 to strongly disagree=1 on a five-point Likert scale (see Appendix K and L). The former had five subsections, namely, emotionality factor , body factor , nervous factor , worry factor , perspiring factor , as identified by its creator; whereas the latter comprised three other subsections: fear of negative evaluation, communication apprehension, and test anxiety. Since the main focus of the present study is speaking and speaking test anxiety in general, the questionnaires were taken as a whole.

To ensure that the items were accurately translated into Turkish, a three-step back-translation process was followed. First, all items were translated into Turkish by the researcher which were then backtranslated into English by another Turkish speaker of English. Then, the two versions in English, namely, the original and the translated versions were compared by a native speaker of English for any missing or misleading expressions.

After the questionnaires had been piloted with a group of students and some items revised, they were administered to the participants. The questionnaires were conducted in Turkish in order to avoid any misinterpretations by the students. The students who did not hand in these questionnaires were removed from the study.

Computer Attitudes Questionnaire

The Computer Attitude Scale (CAS) by Loyd and Gressard (1984), the Computer Attitude Measure (CAM) by Kay (1993), and Computer Familiarity Questionnaire (Kirsch et al., 1998) contributed to the preparation of a 30 item inventory, The Computer Attitudes Inventory (CAI) used in the present study. In addition to the questions created by the researcher, the abovementioned questionnaires were reviewed and appropriate questions were chosen and modified paying attention to the participant profile in the present study in order to collect information regarding the participants' familiarity with and attitudes towards computers. The answers ranged from strongly agree=5 to strongly disagree=1on a five-point Likert scale. The items in the questionnaire focused on perceived self efficacy in using computers, computer anxiety and attitudes towards the internet.

The questionnaire was piloted with a group of university students and all items were found to be adequately comprehensible and clear. Like other questionnaire used in the study, it was conducted in Turkish in order to prevent students from misinterpreting the items. The students who did not submit the questionnaire were not included in the analyses.

Data Collection Procedures

The researcher started by getting permission and asking for assistance from the administration for collecting data from two pre-intermediate and two intermediate classes. Once the classes had been chosen, the instruments used in the study were piloted with students from other classes.

The actual data collection process started with the first speaking tests applied within the same week at both levels. The CASA was administered to one pre-

intermediate and one intermediate class while the other two classes at each level took the FTFsa (see Table 2). The FTFsas were held on the same day at both levels, whereas the CASAs were conducted on two subsequent days due to practical reasons. The students were involved in a brief introduction and practice session just before they took the CASA. Four instructors in total were assigned to administering the two face-to-face tests and rating the computerized tests. Two of the instructors were the speaking class instructors of the classes taking the test and the other two instructors students were unfamiliar with. Briefly, the test takers at each level had two instructors, either for rating their speaking ability (in the CASA) or both administering and rating the test (in the FTFsa). The test takers were given the relevant perceptions inventory, namely, either the FTFsa perceptions questionnaire or the CASA perceptions questionnaire, immediately after they took their test. Both the scoring sheets and the completed questionnaires were collected by the researcher.

On subsequent days, the class instructors were asked to administer the Speaking Test and Speaking Anxiety Questionnaires in addition to the Computer Attitudes Questionnaire (CAQ). Questionnaires were given at different times to avoid participant fatigue. The completed questionnaires were collected by the instructors and handed in to the researcher.

Following a one-month interval, the participants took their second speaking tests. The same procedures as the first tests applied in administration of the second tests and data collection from the second perceptions questionnaires.

Data Analysis

To start with, after scores from all speaking tests gathered, they were put into an excel sheet and an average of the scores from the two raters for each student was determined as the score showing their performance in the speaking test in question. The students who did not attend one of the tests or failed to complete even one of the questionnaires were discarded from the study.

In this study, the researcher utilized the Statistics Package for Social Sciences (SPSS) version 18 to do the quantitative data analysis. The responses to the perceptions questionnaires were analyzed descriptively by looking at the frequency distribution of responses, while the comparisons of various variables were done via non-parametric correlations, Wilcoxon Signed rank test and Three way mixed ANOVA.

Finally, the last part of the perceptions questionnaires containing three open-ended questions was analyzed with qualitative analysis techniques. Responses to each question were categorized into subsections and then interpreted to obtain a summative result. To achieve this, first, all the original responses, which were in Turkish, were typed and grouped under four basic subtitles for each proficiency level: positive attitudes towards the FTFsa, negative attitudes towards the FTFsa, positive attitudes towards the CASA, and negative attitudes towards the CASA. Each response was also coded with the number of the participant who gave the answer. Second, the responses were translated into English by the researcher. Third, the answers with identical meanings were pooled and samples from the groups of responses were chosen randomly to be included in the actual data analysis. Finally, the selection of responses from different levels were interpreted and combined in the data analysis chapter. The responses that correspond to the items in the CASA and

FTFsa perceptions questionnaires were used to support the findings from the descriptive analysis of the questionnaires while the qualitative analysis of the rest of the responses given to the open-ended questions were presented in a separate section in the data analysis chapter.

Conclusion

In this chapter, the setting, the participants, instruments, data collection and analysis procedures were presented. The subsequent chapter will include the findings of the study and the discussion of the findings.

CHAPTER IV: DATA ANALYSIS

Overview of the Study

The aim of this study was to investigate the advantages and disadvantages of computer mediated oral assessment in tertiary level local institutions by comparing a face-to-face speaking assessment (FTFsa) with a computer assisted speaking assessment (CASA). The study focused on students' performance, anxiety levels in the FTFsa and the CASA and their computer attitudes in addition to exploring test takers' perceptions regarding the two modes, at two different proficiency levels. The research questions addressed in the study were:

1. How is the speaking performance of the pre-intermediate and intermediate level test takers at tertiary school affected by the test mode being either the face-to-face (FTFsa) or the computer-assisted speaking assessment (CASA)?
2. What are the students' perceptions of oral assessment?
    a. What are their perceptions of the FTFsa?
    b. What are their perceptions of the CASA?
3. What is the relationship between the anxiety levels of the test takers and the test mode?
    a. What is the relationship between speaking/speaking test anxiety, and test scores?
    b. What is the relationship between speaking/speaking test anxiety, and students' perceptions of FTFsa and CASA?
4. What is the relationship between the computer attitudes of the test takers and the test mode?

   a. What is the relationship between students' computer attitudes and test scores?
   b. What is the relationship between students' computer attitudes and their perceptions of FTFsa or CASA?

5. Depending on the test mode, do the speaking performances, test-mode-related perceptions, and anxiety levels of the test takers at different proficiency levels differ?

The data was gathered at Uludağ University School of Foreign Languages, which offers a preparatory program for learning English as a foreign language before students start their education in their departments. Initially, the participants in the study were 66 students and four instructors. Later, two students were discarded from the study, so a total of 64 students, with 26 students in two pre-intermediate classes and 38 in two intermediate level classes, participated in the study.

The data were collected through two different instruments. The first set of data came from the pre-intermediate and intermediate level FTFsa and CASA tests conducted using a counter-balanced design. The second set of data were individual questionnaires on (1) perceptions of the FTFsa, (2) perceptions of the CASA, (3) speaking anxiety and speaking test anxiety, and (4) computer attitudes.

In this chapter, the analysis of the data is presented in separate sections devoted to the relationships between the abovementioned concepts. First, the scores gained in both test modes will be analyzed quantitatively. Second, the responses to the perceptions questionnaires will be explored through both quantitative and

qualitative methods. Finally, the responses to the anxiety questionnaires will be analyzed by means of quantitative methods.

Inter-Rater Reliability

Table 3

Inter-Rater Reliability Scores

Level	Raters	CASA 1	CASA 2	FTFsa 1	FTFsa 2
Pre-intermediate	1&2	.338	.103	.356	.699*
Intermediate	3&4	.167	.149	.701*	.479*

* Correlation is significant at the 0.01 level (two tailed)

As seen in Table 3 the inter-rater reliability scores were rather low in the CASA at both levels. The only significant correlations were observed between the scores the raters gave in three of the FTFsas at both proficiency levels; yet, the scores given by different raters at the first pre-intermediate FTFsa were not significantly related. This results show that any analyses based on the test scores should be interpreted cautiously.

A Comparison of the Scores

After each of the four classes of students took both the CASA and the FTFsa, their scores were averaged and the averages obtained in the two modes were compared.

Organized according to the order the tests are taken (i.e. group A took the CASA first and group B took the FTFsa first), Table 4 shows the average scores of

students on each test type, the results of Wilcoxon Signed Ranks tests for significant differences between these scores, and the results of Kendall's tau tests of correlation between the tests. Kolmogorov-Smirnov and Shapiro-Wilk tests showed the overall and the pre -intermediate level data not to be normally distributed and the intermediate level data to be normally distributed. Non-parametric tests are used so as to be able to compare the data for the intermediate level with the remaining non-parametric data.

Table 4

Comparison of Test Scores in the CASA and the FTFsa

Level	**G**	**N**	**Mdn**		**IQR**		**Wilcoxon Signed Rank Test**			**Kendall's Tau**	
			C	**F**	**C**	**F**	**T**	**Sig.**	**r**	**ô**	**Sig.**
Overall	A	32	75	79	17.5	29.5	208	.43	-.1	.10	.43
	B	32	77	75	11	19.5	210	.45	-.09	.32*	.01
Pre-int.	A	13	76	92	22	16	0	.002	-.6	.35	.11
	B	13	76	66	11	21	24	.13	-.3	.13	.53
Int.	A	19	74	66	18	22	58	.13	-.2	.06	.69
	B	19	78	80	10	22	76.5	.69	-.06	.36*	.03

Note. G= Group; N= number of participants; C= CASA; F = FTFsa; ô = Kendall's tau correlation coefficient for the FTFsa and CASA scores; r=Effect size; Sig = significance (two-tailed)

As shown in Table 4, the only significant difference was found between the CASA and the FTFsa scores of the pre-intermediate test takers in group A. There was no significant difference between overall CASA scores and overall FTFsa scores. There were also no significant differences between the two tests when the levels were examined separately. This probably means that no matter in which test mode a group of students take the test, they gain similar scores. In other words, the

type of the test has almost no influence on the average scores the groups of test takers obtained.

There was a moderate positive correlation between the overall CASA scores and the overall FTFsa scores, ô = .32, *p(two tailed)* =.01, and intermediate CASA and FTFsa scores, (ô = .36, , *p(two tailed)* =.03) in Group B. No significant correlation was found at the pre-intermediate level in either group, nor were there significant correlations at intermediate level and overall in Group A. Based on the findings, it is possible to say that the FTFsa and the CASA scores give very different rankings to the students, especially at pre-intermediate level, and certainly could not be used in place of each other.

Were Average Scores Affected by Test Type, Level, Doing a Test in the First or Second place?

To investigate further the relationship between scores on the two test types, a three-way mixed ANOVA was computed to see the interaction between the *test type*, the *level* of the test takers and the different *groups taking the tests at different times*. Test scores gained from FTFsa and CASA administrations were taken as the dependent variable, level and group comprised the between-subject variables, and the repeated measures variable was the test mode.

The results revealed that there was no main effect of test mode ($F(1,60)=.52$, $p=.47$), of level ($F(1,60)=.002$, $p=.97$), or of group ($F(1,60)=. 129$, $p=.72$) alone which means that none of these factors changed the average score by themselves.

Since none of these factors changed the average score by themselves it was investigated whether the factors interacted. Figure 6 shows the interaction between

level and test mode. It was found that there is a marginally significant interaction of test mode and level ($F(1,60)=3.87$, $p=.054$). The results indicated that the pre-intermediate students did better on the FTFsa ($M=77.23$, $sd=14.82$) than on the CASA ($M=72.46$, $sd=12.84$), whereas the intermediate students did better on the CASA ($M=76.05$, $sd=10.54$) than on the FTFsa ($M=73.84$, $sd=14.21$). This information indicates that the two levels were affected differently by test mode. However, as we saw above, neither of these differences is statistically significant itself.

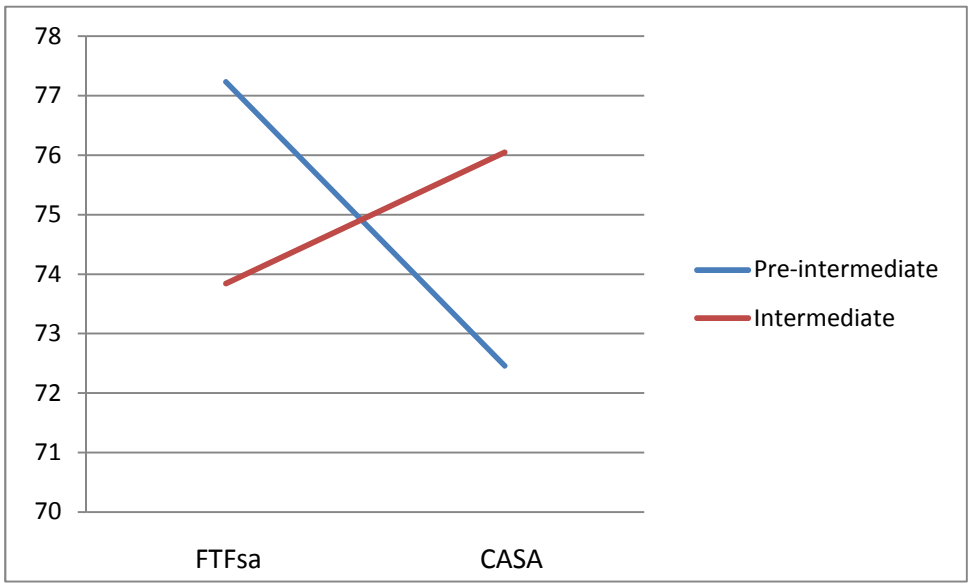

Figure 6 - The interaction between level and test mode

The analysis revealed that there is also a significant interaction of test mode and group ($F(1,60)=4.71$, $p=.034$). As Figure 7 shows, the groups which took the CASA first did better on the FTFsa ($M=75.38$, $sd=15.71$) than on the CASA ($M=72.31$, $sd=13.28$); and the groups which took the FTFsa first did better on the CASA ($M=76.88$, $sd=9.21$) than on the FTFsa ($M=75.06$, $sd=13.30$) independent of their level. The results suggest that both groups did better in their second test than in their first test. This shows that there is a *practice effect* in general and the test takers

improved their speaking practice within the period between the administration of the two tests.

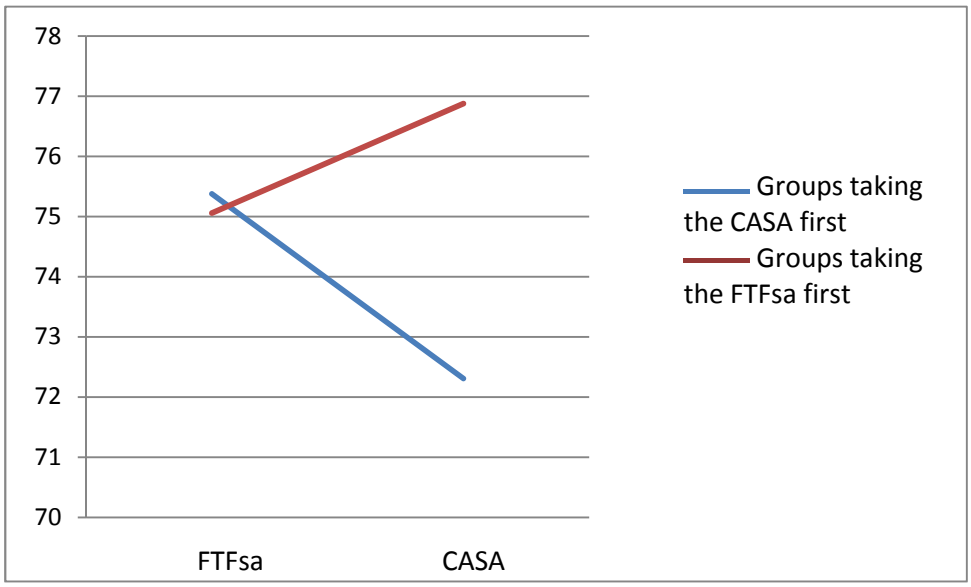

Figure 7 - The interaction between group and test mode

The findings also revealed that there is a significant interaction of mode, level, and group ($F(1,60)=18.04$, $p<.001$). At the pre-intermediate level, the students taking the CASA first did better on the FTFsa ($M=85.69$, $sd=11.6$) than on the CASA ($M=69.54$, $sd=15.94$) whereas the students taking the FTFsa first did better on the CASA ($M=75.38$, $sd=8.42$) than on the FTFsa ($M=68.77$, $sd=12.95$). On the other hand, at the intermediate level, the students taking the CASA first did better on the CASA ($M=74.21$, $sd=11.17$) than on the FTFsa ($M=68.32$, $sd=14.35$) whereas the students taking the FTFsa first did better on the FTFsa ($M=77.89$, $sd=9.81$) than on CASA ($M=79.37$, $sd=12.02$). The findings suggest that the practice effect was actually seen for the pre-intermediate students but not for the intermediate students (see Figures 8 and 9).

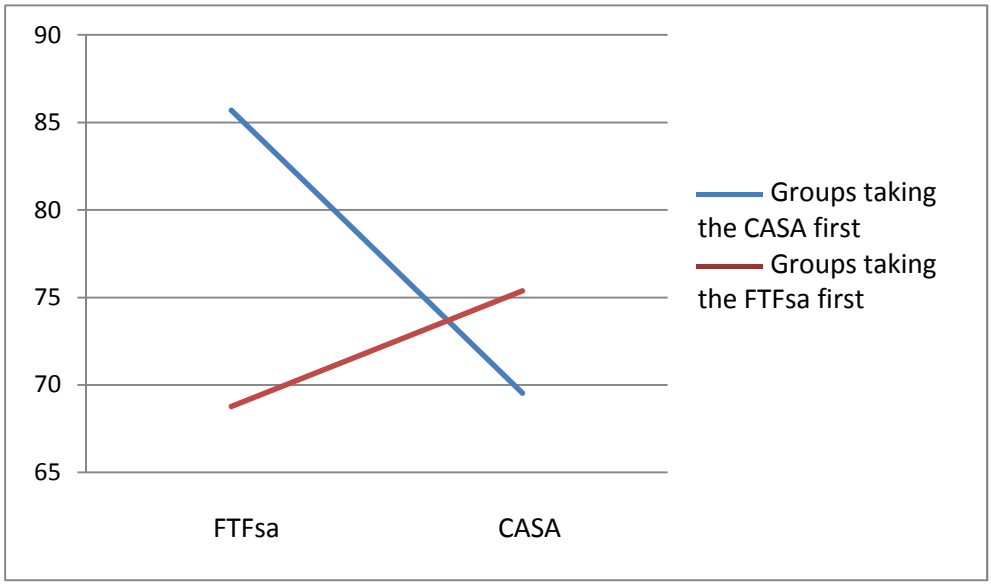

Figure 8 - The interaction between group, test mode, and level (pre-intermediate)

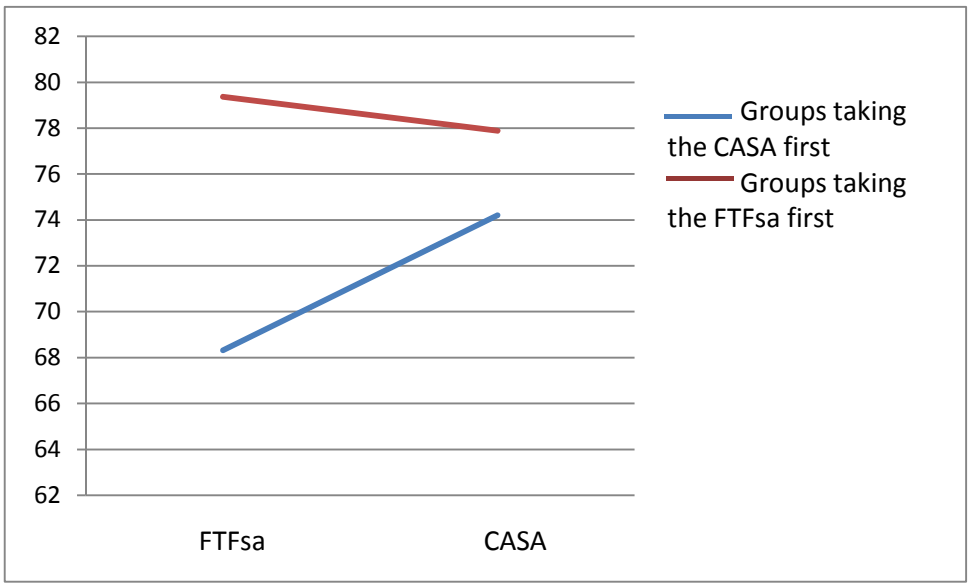

Figure 9 - The interaction between group, test mode, and level (intermediate)

This can be confirmed by checking the difference between first and second tests for the two levels separately.

At the pre-intermediate level, there is a significant difference between first (M=69.15, SE=2.79) and second (M=80.53, SE=2.20) tests (t(25)=4.39, p<.001); however, at the intermediate level, there is no significant differences between first (M=76.79, SE=1.90) and second (M=73.11, SE=2.12) tests (t(37)=1.54, p=.13). Here it is revealed once again that the practice effect existed for the pre-intermediate level only.

The Questionnaires

Initially, all of the questionnaires were piloted with people with similar attributes to the participants of the study and necessary adjustments were made to clarify some items. Once the scores from the tests had been analyzed in detail, the data collected from the four questionnaires were examined. Reliability analyses were run to calculate the Cronbach's Alpha Coefficient for a subscale of the CASA and FTFsa perceptions questionnaires, the speaking test and speaking anxiety questionnaire, and the computer attitudes questionnaire.

In order to answer the second research question, which is about the test takers' perceptions of the FTFsa and the CASA, 66 students were asked to respond to one of the two separate questionnaires after they took each test. Each questionnaire had 29 items designed on a six point Likert scale ranging from "strongly agree = 6" to "strongly disagree = 1" and three open-ended questions. Within the questionnaires, there was a subscale aimed to assess the test-mode-related anxiety levels of the test takers. Some of the items in the questionnaires were reverse coded before initializing the inferential analyses since their meanings were in the opposite direction to the remaining of the questionnaire. A high score on the questionnaire meant positive

feelings towards the test mode in question and low anxiety level in that specific test mode. So, points between 1-2 indicate negative feelings, 3-4 show a neutral perspective, and 5-6 mean that the test taker has positive perceptions of the test mode. Yet, in the descriptive analysis neither the questions nor the scores were reversed. No reliability analysis was run for the overall perceptions questionnaires (see Appendices F, G, H, and I) as they consisted of independent items; however, reliability analyses were computed for the test-mode-related anxiety subscales of the two perceptions scales. Additionally, four pairs or triads of questions measuring similar concepts were used to check for the overall reliability of the two questionnaires. Item 10 was discarded as it decreased the reliability of the questionnaires. Table 5 illustrates the results of the reliability analyses for the test-mode-related anxiety subscales and the above mentioned groups of items.

Table 5

Reliability Analysis of the Perceptions Questionnaires

N of items	**Items/Subscale**	**Cronbach's Alpha**	
		CASA	**FTFsa**
8	Test-mode related anxiety subscale	.83	.81
3	1-13-23	.65	.80
2	8-16	.80	.88
2	17-29	.69	.66
2	20-28	.79	.83

Item by item analysis of the CASA and the FTFsa perceptions Questionnaires

After examining the CASA and the FTFsa perceptions questionnaires via correlations in general, descriptive statistics were computed in order to investigate the frequency distribution of the answers of the participants for each item of the questionnaires both overall and at separate levels. The six scale categories were

collapsed into three new categories; first, “strongly agree” and “agree”; second, “partly agree” and “partly disagree”; and third, “disagree” and “strongly disagree”. With related items grouped together, descriptive statistics for all items were summarized overall in six tables and for separate levels in another group of six tables, and the results were interpreted to show the details of the test takers’ perceptions of the two test modes. Qualitatively analysed and interpreted, the responses to the open ended questions asked at the end of the questionnaires were added to the descriptive analysis results to support the findings when appropriate (For a full list of open-ended responses in Turkish see Appendix J). The results for the test-mode related anxiety subscales in both questionnaires are shown in Table 6 for all participants and Table 7 for the participants at different proficiency levels.

The averages of the percentages of responses given to all questions in the anxiety subscales showed that 26.9% of all participants agreed or strongly agreed that they were anxious in the FTFsa. An even higher number of participants (46.8%) reported being anxious in the CASA.

Items 2, 3, 4, and 6 in Table 6 show that the test takers were anxious prior to and during both tests and after the CASA, but their anxiety levels were relatively higher for the CASA in all situations. For instance, noticeably, 34.9% of the test takers stated that they were anxious after the CASA while 13.7% stated that they were anxious after the FTFsa.

When the pre-intermediate and intermediate levels are examined separately (Table 7), it is seen that there are differences between levels at some points. As the responses to item 2 indicate, more than 70% of the test takers at both levels reported

being more or less anxious before the FTFsa. At the intermediate level, even more test takers felt anxious and tense before the CASA. Surprisingly, though, 39.3% of the pre-intermediate test takers stated that they were not anxious before the CASA. Obviously, however, their anxiety levels changed during the speaking tests as responses to item 3 suggest.

Table 6

Test-Mode-Related Anxiety Subscale (Overall)

	R	**Test mode**	**N**	**SA/A %**	**PA/PD %**	**D/SD %**	**Mean**	**sd**
2. I felt rather tense and anxious before the speaking test.	R	F	66	39.4	36.4	24.3	3.88	1.62
		C	66	47	27.3	25.7	3.88	1.66
3. I felt tense and anxious during the speaking test.	R	F	66	30.3	34.9	34.9	3.56	1.61
		C	66	54.5	28.8	16.6	4.30	1.44
4. I felt tense and anxious after the speaking test.	R	F	66	13.7	27.3	59.1	2.58	1.53
		C	66	34.9	22.7	42.4	3.44	1.71
6. I felt very relaxed before the speaking test.		F	66	15.2	36.4	48.4	2.86	1.55
		C	66	12.1	25.8	62.1	2.53	1.44
7. I was very afraid of making mistakes during the speaking test.	R	F	66	27.3	36.4	36.3	3.55	1.44
		C	65	47	24.2	27.3	4.02	1.68
14. It relieved me to see someone listening to me during the speaking test.		F	66	33.3	48.5	18.2	3.85	1.42
14. It relieved me to see that no one was listening to me during the speaking test.		C	66	19.7	31.8	48.5	2.88	1.63
18. The fact that our speaking will be tested motivates me in terms of speaking English.		F	66	36.4	34.9	28.8	3.64	1.53
		C	66	18.2	30.3	51.5	2.80	1.44
19. The speaking test helped to decrease my fears about speaking English.		F	66	30.3	42.5	27.3	3.55	1.40
		C	66	9.1	36.4	54.6	2.65	1.38
10. It was irritating that I couldn't ask for clarification from the test giver.*		F	63	9.1	16.7	69.7	2.35	1.24
		C	66	21.2	33.3	45.4	2.95	1.66

Table 7

Test-Mode-Related Anxiety Subscale for Different Levels

	R	Level	Test mode	N	SA/A %	PA/PD	D/SD %	Mean	Sd
2. I felt rather tense and anxious before the speaking test.	R	P	F	28	35.7	39.3	25	3.79	1.66
		P	C	28	42.8	17.9	39.3	3.57	1.77
		I	F	38	42.2	34.3	23.7	3.95	1.61
		I	C	38	50	34.3	15.8	4.11	1.55
3. I felt tense and anxious during the speaking test.	R	P	F	28	35.8	25	39.3	3.68	1.58
		P	C	28	53.6	25	21.5	4.18	1.44
		I	F	38	26.3	42.2	31.6	3.47	1.65
		I	C	38	55.2	31.6	13.1	4.39	1.46
4. I felt tense and anxious after the speaking test.	R	P	F	28	10.7	28.6	60.7	2.54	1.40
		P	C	28	28.5	25	46.5	3.32	1.49
		I	F	38	15.8	26.3	57.9	2.61	1.63
		I	C	38	39.5	21	39.5	3.53	1.87
6. I felt very relaxed before the speaking test.		P	F	28	17.8	42.9	39.3	2.96	1.57
		P	C	28	14.3	17.9	67.9	2.54	1.47
		I	F	38	13.2	31.6	55.3	2.79	1.56
		I	C	38	10.5	31.6	57.9	2.53	1.44
7. I was very afraid of making mistakes during the speaking test.	R	P	F	28	21.4	42.8	35.7	3.50	1.37
		P	C	27	35.7	25	35.7	3.63	1.77
		I	F	38	31.6	31.6	36.8	3.58	1.51
		I	C	38	55.2	23.7	21	4.29	1.57
14. It relieved me to see/ not to see someone listening to me during the speaking test.		P	F	28	46.4	39.3	14.2	4.21	1.52
		P	C	28	17.8	25	57.1	2.64	1.68
		I	F	38	23.7	55.3	21.1	3.58	1.30
		I	C	38	21.1	36.9	42.1	3.05	1.59
18. The fact that our speaking will be tested motivates me in terms of speaking English.		P	F	28	53.6	25	21.4	4.00	1.56
		P	C	28	21.4	32.2	46.4	3.11	1.44
		I	F	38	23.7	42.1	34.2	3.37	1.47
		I	C	38	15.8	28.9	55.2	2.58	1.42
19. The speaking test helped to decrease my fears about speaking English.		P	F	28	42.8	32.1	25	3.89	1.49
		P	C	28	14.3	39.2	46.4	2.93	1.43
		I	F	38	21.1	50	28.9	3.29	1.29
		I	C	38	5.3	34.2	60.5	2.45	1.32
10. It was irritating that I couldn't ask for clarification from the test giver.*		P	F	27	7.2	17.8	71.4	2.19	1.30
		P	C	28	35.7	21.4	42.9	3.50	1.91
		I	F	36	10.5	15.8	68.4	2.47	1.20
		I	C	38	10.5	42.1	47.3	2.55	1.35

Note. I= Intermediate, P= Pre-intermediate, R= Reversed Item, F= FTFsa, C= CASA N= Number of respondents,*= Item was discarded while calculating inferential statistics.

Both the pre-intermediate and the intermediate level test takers felt anxious during the CASA, with a small number of students as exceptions. During the FTFsa, however, one third of the pre-intermediate students were anxious while another one third were quite relaxed. Similarly, almost one third of the intermediate level test takers felt at ease, but the number of intermediate test takers who responded as "undecided" exceeded the number of those at ease. Answers given to item 4 display that a good number of test takers at both levels (60.7%, 46.5%, 57.9%, 39.5 % respectively) were not tense or anxious after either speaking test. Yet, 28.5% of the pre-intermediate and 39.5% of the intermediate participants, which is indeed equal to the number of intermediate level participants with no anxiety, reported that they were still anxious and tense after the CASA. Responses to item 6 revealed that the participants at either level were not really relaxed before either test mode. Although being anxious before or during a test might be considered normal, having a high level of anxiety even after the test may indicate that the students were actually anxious or nervous because they faced something new, which means extra challenge.

As seen in all test takers' response rates to item 7 in Table 6, more than one third of the tests takers who took the FTFsa felt neutral about making mistakes, but still another one third of them noted they were not afraid of making mistakes during the test. On the other hand, almost half of them seem to have felt afraid of making mistakes in the CASA. This may have resulted from interlocutor interference in the FTFsa given that interlocutors typically try to relieve test takers during the interviews. The responses given to the open ended questions, which will be discussed in detail later in the chapter, also confirm the possibility of this tendency.

The responses to item 14 point out that nearly half of the students were apathetic to the fact that they were being listened to by someone in the FTFsa, yet 33.3% seems to have appreciated it while 18% did not. On the other hand, in the CASA, 48.5% were uncomfortable with the fact that no one was actually listening to them at the moment they were taking the test. At different levels, the findings were similar. The results from the qualitative analysis of the open-ended questions confirm that the test takers tended to value the existence of a live interlocutor listening to them. To start with the pre-intermediate level test takers' perceptions of the FTFsa, it was seen that most of the participants reported having felt relieved and comfortable thanks to the positive attitudes of the interlocutors during the face-to-face test. For instance, participant 7 wrote:

> *Nothing made me feel uncomfortable in the test. The interlocutors were very understanding so I felt very comfortable during the test.*

Similarly, participant 9 emphasized the helpfulness of the interlocutors by noting:

> *I liked that my teachers were understanding. The fact that they were smiling and showed that I was being listened to helped me feel really confident.*

Participants 2 and 5 also indicated that they liked that the interlocutors tried to relieve them or reduce their stress, and participant 6 stated that the interlocutors had much better attitudes than normal and they were very patient. Participants 3 and 12 implied that the interlocutors seemed to have understood them even though their performance was not very good, which also helped them feel less anxious.

It seems that the pre-intermediate level participants were also attracted to the fact that the test givers listened and paid attention to them as they were talking. Participant 4 noted that he liked that the interlocutors listened to him carefully, as did participant 25. Similarly, participant 13 wrote:

It was really good that someone actually listened to me. At least it relieved me a little. And I liked it more as I continued to speak.

The existence of someone listening to the test takers and the attitudes of the interlocutors were among the most noticeable points the test takers liked about the FTFsa at the intermediate level, too. Numerous test takers reported them to be understanding, friendly, smiling, motivating, and relaxing. The sample statements below show what the test takers at intermediate level thought about the interlocutors and their influence on the test:

Participant 31: The attitudes of the teacher were very comforting. It was more like having a chat instead of taking a test.

Participant 41: The encouraging attitudes and friendly personalities of the test givers were the best aspects of the test for me. I think a face-to-face test is much more effective than a computerized one. The test givers remind the test takers of some of the structures to be used and help them. Moreover, there is a more convenient atmosphere.

Participant 51: The teachers did not push us very hard to speak.

As is obvious from the lines above, some test takers at both levels thought the interlocutors motivated and calmed them down. However, as stated by participant 41, they might have interfered more than needed and helped some students answer some of the questions, which would have decreased the reliability of the test. Moreover, one participant (26) stated that being listened to by an unfamiliar tester bothered him and he could have performed better if his speaking instructor had been the only one testing him. Surprisingly, participant 15 was worried because someone was listening to her at all. Participant 2 did not like her speech to be scored by someone while she was still speaking. The excerpts below also confirm that some interlocutor behaviours might have triggered the tenseness at the intermediate level:

> Participant 33: The test was conducted together with two interlocutors. I was disturbed because one was taking notes while the other one was constantly asking questions.
>
> Participant 36: During the conversations that we had with two interlocutors, one of them talked fluently as if we were chatting whereas the other was only watching and taking notes, which was irritating.

Briefly, these responses indicate that there were test takers, though few, who felt uncomfortable in the FTFsa despite the efforts of the interlocutors, too. On the other hand, confirming the findings from item 14, some test takers at the pre-intermediate level can be said to have felt relaxed during the CASA as exemplified below:

> Participant 2: Talking to a computer instead of a human helped me feel relieved.
>
> Participant 15: It was nice to know that no one was listening to me.
>
> Participant 22: The fact that everybody was engaged in their own test helped me feel more comfortable.
>
> Participant 25: I didn't have difficulty in understanding the questions and I easily responded to them.

The excerpts below display that some participants at the intermediate level also found it more convenient to talk to a computer:

> Participant 47: Taking the test in a computerized environment relieved me.
>
> Participant 52: It was more comfortable to talk to the computer instead of a teacher.

In short, the presence of a live interlocutor in the FTFsa was appreaciated, yet some participants found it more convenient to talk to a computer alone. The problems as associated with lack of an on-site test giver in the CASA were more prominent, though. The test takers thought it affected their performance negatively and it was irritating to talk to a computer instead of a person. The excerpts below show that many students disliked the lack of interaction:

> Participant 7: I was uncomfortable with the lack of communication. I think the tests should be face-to-face. I felt uncomfortable in the computer-assisted test.
>
> Participant 13: The fact that nobody was listening to me and I couldn't remember things disturbed me. I think the speaking tests shouldn't be done on the computer.
>
> Participant 17: It was inconvenient to talk to the computer in the test. Perhaps the words were not understood clearly…
>
> Participant 19: Having a computer in front of me did not create an intimate atmosphere at all.
>
> Participant 26: I don't think I was able to fully express my ideas since there was no one listening to me.

Fewer intermediate level test takers focused on the lack of a live interlocutor: Participant 66 reported that she found it bothering to speak by herself and record her voice while some others compared it with the FTFsa, finding the latter more comfortable (participant 35), and criticizing the CASA by saying it was irritating not to have someone listening to you (participant 46). Participant 63 thought responding to a computer was rather annoying because he couldn't make sure if his responses were recorded. In short, all of these participants seem to have disliked the lack of interaction and a live interlocutor to talk to.

When examined overall, answers given to item 18 show that taking the FTFsa raised positive feelings towards speaking English in 36.4% of respondents. Nevertheless, a considerably high number of students (28.8%) thought the opposite way. As for the CASA, more than half responded that the existence of speaking tests did not motivate them, so obviously they had negative feelings towards the CASA. The results obtained from item 19 also confirm these findings. The results were dissimilar at different levels, though. Only the pre-intermediate FTFsa motivated the students and helped them gain some self confidence in speaking English, as seen in Table 7.

Item 10 focused on the perceptions of the test takers of the interaction in the speaking tests. Surprisingly, neither of the tests seems to have irritated the participants due to lack of clarification requests from the test givers. However, when levels were investigated separately, a considerable number of pre-intermediate test takers (35.7%) were found to be annoyed by the fact that they were unable to ask for clarification in the CASA.

In brief, the participants at both levels seem to have experienced more tension in the CASA although there were slight differences at some points.

Tables 8 and 9 display the answers to the three questions about the perceived difficulty of the speaking tests for the participants in general and at different levels.

Table 8

Questions about the Difficulty of the Speaking Tests (Overall)

	R	**Test mode**	**N**	**SA/A %**	**PA/PD %**	**D/SD %**	**Mean**	**sd**
1. The speaking test was very difficult.	R	F	66	9.1	34.9	56.1	2.76	1.27
		C	66	36.3	45.5	18.2	4.12	1.42
13. I don't think that I can get a good mark from the speaking test.	R	F	66	22.7	48.5	28.7	3.52	1.38
		C	66	59	21.2	19.7	4.38	1.68
23. I think I can get a good mark from the speaking test.		F	66	28.8	48.5	22.7	3.65	1.42
		C	66	9.1	19.7	71.2	2.24	1.37

Note. R= Reversed Item, F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation

Table 9

Questions about the Difficulty of the Speaking Tests as Perceived at Different Levels

	R	Level	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	sd
1. The speaking test was very difficult.	R	P	F	28	7.1	35.8	57.2	2.68	1.33
		P	C	28	32.1	42.8	25	3.96	1.45
		I	F	38	10.6	34.2	55.3	2.82	1.24
		I	C	38	39.5	47.4	13.2	4.24	1.40
13. I don't think that I can get a good mark from the speaking test.	R	P	F	28	21.4	46.5	32.1	3.43	1.45
		P	C	28	75	10.7	14.3	4.96	1.55
		I	F	38	23.7	50	26.3	3.58	1.34
		I	C	38	47.3	28.9	23.7	3.95	1.67
23. I think I can get a good mark from the speaking test.		P	F	28	32.1	46.4	21.4	3.82	1.41
		P	C	28	10.7	7.1	82.2	1.93	1.35
		I	F	38	26.4	50	23.7	3.53	1.42
		I	C	38	7.9	28.9	63.1	2.47	1.35

Note. R= Reversed Item, I= Intermediate, P= Pre-intermediate, F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation

As seen in Table 8, item 1 directly asks about the difficulty of the tests and only around 10% of the test takers thought the FTFsa was a difficult test and more than half completely disagreed. On the other hand, about one third of the test takers found the CASA difficult and only 18% disagreed. Briefly, the test takers at both pre-intermediate and intermediate levels found the CASA difficult while fewer participants thought the same for the FTFsa.

Items 13 and 23 focus on test takers' expectations about the grade they would get on each test mode. The results reveal that 48.5% of participants were uncertain about the scores they would get, but still a lot (28%) thought they would get high ones on the FTFsa. As for the CASA, considerably high numbers of participants believed that they would get low scores. In brief, the majority of the participants at

both levels found the CASA difficult but it was the opposite for the FTFsa. It is important to mention that both tests were identical in terms of content, style and the number of questions.

Tables 10 and 11 show the results for the questions looking for how the participants thought the speaking tests would influence their classroom attendance.

Although the same questions were asked in both test modes, students' perceptions in terms of how the test would affect their attendance seem to differ. While around 43% of test takers only partly agreed that taking the FTFsa would increase their attendance to related classes, more than 40% thought that taking the CASA would not change their attendance rates. As for separate levels, although the FTFsa was promising in terms of increasing the attendance to classes at the pre-intermediate level, neither test mode seems to have an effect on the intermediate level students' attendance. As the questions were identical, it is open for investigation why students perceived the CASA so differently and why they might have thought that it had no relation to what was introduced in the speaking classes.

Table 10

Questions about the Relationship between Speaking Tests and Classroom Attendance

	R	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	sd
8. The speaking test will increase my attendance to speaking classes.		F	66	24.2	43.9	31.8	3.36	1.41
		C	65	21.2	33.3	44	3.05	1.75
16. The speaking test will increase my attendance to the classes where speaking is practiced.		F	66	30.3	42.4	27.3	3.59	1.38
		C	66	16.6	36.3	47	2.86	1.53

Note. F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation

Table 11

Questions about the Relationship between Speaking Tests and Classroom Attendance at Different Levels

	R	Level	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	Sd
8. The speaking test will increase my attendance to speaking classes.		P	F	28	46.4	32.1	21.4	3.86	1.35
		P	C	27	28.6	32.1	35.8	3.52	1.86
		I	F	38	7.9	52.7	39.5	3.00	1.35
		I	C	38	15.8	34.2	50	2.71	1.60
16. The speaking test will increase my attendance to the classes where speaking is practiced.		P	F	28	57.2	28.6	14.3	4.18	1.24
		P	C	28	25	32.1	42.9	3.18	1.61
		I	F	38	10.5	52.6	36.9	3.16	1.32
		I	C	38	10.5	39.5	50	2.63	1.46

Note. R= Reversed Item, I= Intermediate, P= Pre-intermediate, F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation

Tables 12 and 13 below include the frequency distribution of the answers to the items questioning whether the speaking tests were perceived as good tools for assessing the participants' speaking abilities.

Table 12
Questions about the Quality of the Tests (overall)

R	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	sd
5.The speaking test effectively tested what was taught in speaking classes or in the speaking sections of other classes.	F	66	40.9	45.4	13.6	4.11	1.31
	C	64	34.8	36.4	25.8	3.67	1.56
15. The speaking test helped me fully reflect my speaking ability.	F	66	16.7	48.4	34.8	3.23	1.33
	C	65	6.1	22.7	69.6	2.26	1.39
17. The speaking test was a good tool for me to show my speaking ability.	F	66	30,3	43.9	25.8	3.61	1.41
	C	65	9.1	40.9	48.5	2.78	1.36
29. The speaking test allowed me to show my strong and weak points in speaking English.	F	66	40.9	37.9	21.3	3.91	1.37
	C	66	22.8	50	27.3	3.39	1.47

Note. F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation

Seeking the participants' perceptions of the face validity of the speaking tests, item 5 inquired whether the speaking tests effectively tested what was taught in speaking classes. Forty percent of the test takers thought the FTFsa did it successfully, and 34% found the CASA to be effective in this respect. In both cases, a considerable number of participants partly agreed that the tests were effective. This suggests that the test takers favored the FTFsa over the CASA again, and considered it to be a more functional tool or good choice for testing speaking ability gained

through the speaking classes when all participants are considered. As for different levels, most pre-intermediate level participants thought both the CASA and the FTFsa assessed what was taught in the speaking classes effectively. For the intermediate level, the results were less clear as a good number of participants (57.9 % and 36.9%) were undecided about the issue mentioned, yet it is clear from the mean scores that they are likely to have found the FTFsa effective in this respect but not the CASA, unlike their pre-intermediate peers.

Responses to items 15 and 17 show that only 6.1% and 9.1% of the participants thought that the CASA let them show their full speaking ability and it was a good tool respectively, while the numbers were higher for the FTFsa (16.7% and 30.3%). Apparently, the participants, in general, believed that the FTFsa was more likely to reflect their oral proficiency although there were also a lot of participants who believed FTFsa was not good, either. As shown in Table 13, 25% of the participants at the pre-intermediate level thought that the FTFsa helped them to fully demonstrate their speaking ability and 42.8% thought it was a good tool for testing their speaking ability in English; however, it was the opposite for the CASA for most of them. As for the intermediate level participants, none of the tests were appealing with respect to their quality in general.

Table 13

Questions about the Quality of the Tests at Different Levels

	R	Level	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	sd
5.The speaking test effectively tested what was taught in speaking classes or in the speaking sections of other classes.		P	F	28	60.7	28.5	10.7	4.50	1.42
		P	C	27	42.8	35.7	17.9	4.15	1.48
		I	F	38	26.4	57.9	15.8	3.82	1.15
		I	C	37	29	36.9	31.6	3.32	1.54
15. The speaking test helped me fully reflect my speaking ability.		P	F	28	25	57.2	17.8	3.71	1.24
		P	C	27	7.1	21.4	67.9	2.37	1.44
		I	F	28	10.5	42.1	47.4	2.87	1.29
		I	C	38	5.3	23.7	71.1	2.18	1.37
17. The speaking test was a good tool for me to show my speaking ability.		P	F	28	42.8	39.2	17.9	4.04	1.29
		P	C	28	17.9	42.8	39.3	3.21	1.39
		I	F	38	21	47.4	31.6	3.29	1.43
		I	C	37	2.6	39.5	55.2	2.46	1.26
29. The speaking test allowed me to show my strong and weak points in speaking English.		P	F	28	53.6	35.7	10.7	4.25	1.37
		P	C	28	35.7	42.9	21.4	3.86	1.48
		I	F	38	31.5	39.5	29	3.66	1.34
		I	C	38	10.5	39.5	50	3.05	1.39

Note. R= Reversed Item, I= Intermediate, P= Pre-intermediate, F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation

Responses to item 29 differ from those of 15 and 17 in frequency distribution. It is seen that a considerably high number of students (40.9%) thought that the FTFsa helped them show their strengths and weaknesses. A great number of participants (50%) partly agreed with this statement when it came to the CASA. On the other hand, while around 23% thought the CASA was successful in showing their strengths and weaknesses, 27% totally disagreed. Briefly, most of the participants thought that both tests helped them see their strength and weaknesses, yet the FTFsa was found

useful by more participants. However, examining the two levels separately revealed that all but the intermediate level CASA were seen as letting the participants at both levels demonstrate their strengths and weaknesses in speaking English, which suggests that they were not satisfied with the content and style of the questions in the intermediate CASA.

Similarly, according to the responses given to the open-ended questions, the pre-intermediate students seem to have liked about the FTFsa is the content, structure, and presentation of the questions. Participant 8 thought the questions were easy and appropriate for the speaking proficiency level and participant 15 stated the questions were clear. Participant 26 thought it was nice to proceed in a planned way with the help of the papers that give them an outline for the longer presentation-type questions and the use of visuals also aided their speech. Obviously they were also content with the content and comprehensiveness of the questions as can be understood from the statements below:

> Participant 21: The questions included conversations we could have in daily life. Moreover, the topics we focused on and practiced a lot in class were questioned.
>
> Participant 28: The topic options were a lot so we had the chance to show different aspects of our speaking ability.

Similarly, some students at the intermediate level also mentioned their positive thoughts about the content and style of the questions in the FTFsa. For instance, participant 34 said she liked that the questions were from their curriculum. Participant 33 took it from another perspective and noted that she liked the flow of the questions which started with daily life, then focused on more specific details, and

then assessed what they learnt in the speaking class. Some intermediate level participants also liked the individualized nature and authentic topics of the FTFsa.

The content and the format of the questions in the CASA appealed to some students as well. Participants 4 and 6 at the pre-intermediate level stated that they liked the questions and that they were presented in written format. As noted by participants 22 and 28, the questions were thought to include samples from daily speech, and the level of language was moderate. The test was said to help gain self confidence due to these factors. On the other hand, none of the participants at the intermediate level made similar comments.

Obviouly, although there were test takers who appreaciated the CASA questions to some extent, the FTFsa was preffered over the CASA in terms of content and format of the questions.

Tables 14 and 15 present the frequencies of the responses given to questions 20 and 28. The items inquire whether the tests were perceived as comprehensive in terms of quality and quantity.

As indicated in Table 14, 36.4% of the participants found the FTFsa comprehensive enough whereas 27.3% found the CASA to be so. When the perceptions regarding the number and variety of questions in the tests were questioned, a large number of participants seem to have preferred to be neutral. Nevertheless, there were still more participants content with the variety and amount of the questions in the FTFsa than in the CASA.

Table 14

Questions about the Comprehensiveness of the Tests

	R	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	sd
20. The speaking test was comprehensive enough.		F	66	36.4	42.5	21.2	3.82	1.30
		C	66	27.3	36.3	36.4	3.32	1.56
28. There were adequate amount and variety of questions in the speaking test to test my speaking ability.		F	66	22.7	45.5	31.9	3.39	1.28
		C	66	15.1	42.4	42.4	2.92	1.39

Note. F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation

Table 15

Questions about the Comprehensiveness of the Tests at Different Levels

	R	Level	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	sd
20. The speaking test was comprehensive enough.		P	F	28	53.6	42.8	3.6	4.43	1.06
		P	C	28	39.3	39.3	21.4	3.82	1.51
		I	F	38	23.7	42.1	34.2	3.37	1.28
		I	C	38	18.4	34.2	47.4	2.95	1.50
28. There were adequate amount and variety of questions in the speaking test to test my speaking ability.		P	F	28	28.6	53.6	17.9	3.75	1.20
		P	C	28	21.4	46.4	32.2	3.29	1.38
		I	F	38	18.4	39.5	42.1	3.13	1.29
		I	C	38	10.5	39.5	50	2.66	1.36

Note. R= Reversed Item, I= Intermediate, P= Pre-intermediate, F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation

As for different levels, more than half of the pre-intermediate level participants found the FTFsa to be adequately comprehensive, yet the CASA was found comprehensive by a good number of participants (39.3%), too. Interestingly, though, a considerable number of pre-intermediate level participants responded that the CASA included an inadequate variety of questions to test their speaking ability, which may have resulted from their general negative attitudes towards the CASA because the questions were exactly the same in both test modes. The intermediate level participants, on the other hand, were mainly dissatisfied with the comprehensiveness of both tests and the diversity of questions.

Finally, Table 16 illustrates the frequency distribution of the responses for the remaining items in the CASA and the FTFsa perceptions questionnaires. Table 17 gives the same results for separate levels.

Item 9, inquiring whether the tests represented real life experiences, yielded opposite results for the two test modes. While 42.4% of the test takers confirmed that speaking to a test giver in the FTFsa represented a real life experience, merely 12.1% responded that speaking to the computer represented a real life speaking experience, which is an anticipated finding. The results were similar when different levels were examined separately.

Table 16

Individual Questions that do not Belong to a Specific Category

	R	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	sd
9. The fact that I responded to a test giver did not represent a real-life speaking experience I can have.	R	F	65	18.2	37.9	42.4	3.06	1.45
9. The fact that I responded to a computer did not represent a real-life speaking experience I can have.		C	66	54.5	33.3	12.1	4.50	1.43
11. The amount of instructions given during the speaking test was too much.	R	F	66	7.6	30.3	62.1	2.55	1.13
		C	66	10.6	41	48.5	2.80	1.28
12. I could flexibly respond to the questions asked in the speaking test.		F	66	31.8	51.5	16.7	3.82	1.30
		C	66	15.1	42.4	42.4	2.98	1.35
21. There was no interaction during the speaking test.	R	F	66	3	27.3	69.7	2.27	1.03
		C	65	30.3	47	21.2	3.71	1.44
22. I would like to have my speaking tests in face-to-face /computerized format from now on.		F	65	59	30.3	9.1	4.46	1.40
		C	65	13.7	28.8	56	2.57	1.58
24. Adequate time was given to answer each question in the speaking test.		F	66	78.8	13.6	7.6	4.85	1.15
		C	66	45.5	28.8	25.7	3.83	1.66
25. The visual support materials helped me answer the questions.		F	66	48.5	45.5	6	4.35	1.19
		C	66	36.4	33.3	30.3	3.56	1.53
26. I could easily organize my thoughts in the speaking test.		F	66	16.6	51.5	31.9	3.29	1.25
		C	66	10.6	31.9	57.6	2.55	1.41
27. I think the speaking test is not a fair one.	R	F	66	15.1	21.3	63.7	2.68	1.44
		C	65	28.8	39.4	30.3	3.60	1.58

Note. R= Reversed Item, F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation.

Table 17

Responses to Individual Questions that do not Belong to a Specific Category at Different Levels

	R	Level	Test mode	N	SA/A %	PA/PD %	D/SD %	Mean	sd
9. The fact that I responded to a test giver did not represent a real-life speaking experience I can have.	R	P	F	28	14.3	39.3	46.4	2.82	1.44
		P	C	28	60.7	25	14.3	4.61	1.47
		I	F	37	21	36.9	39.5	3.24	1.46
9. The fact that I responded to a computer did not represent a real-life speaking experience I can have.		I	C	38	50	39.5	10.5	4.42	1.42
11. The amount of instructions given during the speaking test was too much.	R	P	F	28	14.3	21.4	64.3	2.54	1.29
		P	C	28	7.1	32.1	60.7	2.50	1.10
		I	F	38	2.6	36.9	60.5	2.55	1.32
		I	C	38	13.2	47.3	39.4	3.03	1.36
12. I could flexibly respond to the questions asked in the speaking test.		P	F	28	28.6	57.1	14.3	3.79	1.22
		P	C	28	21.4	28.5	50	2.96	1.47
		I	F	38	34.2	47.4	18.4	3.84	1.36
		I	C	38	10.5	52.6	36.9	3.00	1.27
21. There was no interaction during the speaking test.	R	P	F	28	7.2	28.6	64.3	2.39	1.34
		P	C	28	39.3	35.7	25	3.82	1.63
		I	F	38	0	26.4	73.7	2.18	0.73
		I	C	37	23.7	55.3	18.4	3.62	1.29
22. I would like to have my speaking tests in face-to-face format from now on.		P	F	27	75	21.4	0	4.93	0.99
		P	C	27	7.1	32.2	57.1	2.41	1.33
22. I would like to have my speaking tests in computerized format from now on.		I	F	38	47.4	36.9	15.8	4.13	1.56
		I	C	38	18.4	26.4	55.2	2.68	1.74
24. Adequate time was given to answer each question in the speaking test.		P	F	28	75	17.8	7.2	4.75	1.23
		P	C	28	50	14.3	35.8	3.82	1.88
		I	F	38	81.5	10.5	7.9	4.92	1.10
		I	C	38	42.1	39.5	18.4	3.84	1.49
25. The visual support materials helped me answer the questions.		P	F	28	67.8	25	7.2	4.75	1.29
		P	C	28	42.8	25	32.1	3.75	1.62
		I	F	38	34.2	60.6	5.2	4.05	1.03
		I	C	38	31.5	39.5	29	3.42	1.46
26. I could easily organize		P	F	28	21.5	50	28.5	3.43	1.31

my thoughts in the speaking test.		P	C	28	10.7	21.4	67.8	2.32	1.44
		I	F	38	13.1	52.6	34.2	3.18	1.20
		I	C	38	10.5	39.5	50	2.71	1.39
27. I think the speaking test is not a fair one.	R	P	F	28	10.7	14.3	81.3	2.32	1.49
		P	C	28	21.5	32.1	46.4	3.14	1.77
		I	F	38	18.4	26.3	55.3	2.95	1.37
		I	C	37	34.2	44.7	18.4	3.95	1.35

Note. R= Reversed Item, I= Intermediate, P= Pre-intermediate, F= Face-to-Face speaking assessment, C= Computer assisted speaking assessment, N= Number of respondents, SA/A= Strongly agree/agree, PA/PD= Partly agree/partly disagree, D/SD= Disagree/strongly disagree, sd= Standard deviation.

Item 11 asked if the amount of instructions given during either test was excessive. As the both overall and level-specific responses disclosed, a noticeably high number of participants thought the instructions were not too much in either test mode, although the overall number of participants favoring the FTFsa outreached those favoring the CASA by 13.6 percent. This suggests that most of them were satisfied with the instructions given in the tests, which adds to the reliability of the tests as it probably means that the instructions were clear and adequate for each student to understand the questions.

According to the responses to item 12, 31.8% noted that they could flexibly answer the questions in the FTFsa, while only half as many of them thought they could do the same in the CASA. The results were similar for separate levels, yet more students at the pre-intermediate level than intermediate disliked the CASA in this respect. The intermediate level participants might have performed more comfortably in the CASA due to their higher level of proficiency.

In Table 16, it is seen that most participants thought that there was some interaction in the FTFsa, while for the CASA some thought there was no interaction

at all and others believed the opposite. Yet, most of them were undecided. So it seems that how they perceived the word “interaction” varied, or some felt like they were actually responding to someone as they were talking to the computer.

Item 22 questions which test mode the test takers would prefer. Obviously, the majority would prefer the FTFsa rather than the CASA, which was said to be preferred by only 13.7% overall. At the pre-intermediate level, 75 % clearly stated that they wanted to take face-to-face speaking tests later on, while the proportion was 47.4 % at the intermediate level.

According to results from item 24, the time given in both test modes was found adequate. The overall number of the participants thinking that enough time was given in the FTFsa outreached the CASA number by 33.3 percent, though. Few participants (6% for the FTFsa, 30% for the CASA) found the visual support materials unhelpful, as responses to item 25 indicate. Items 24 and 25 also revealed that the pre-intermediate students were mostly content with the time and visuals materials provided in both the CASA and the FTFsa, but still there were more than 30% who found the time inadequate and the visuals unsupportive in the CASA. Although the reason for such a distribution in answers is not clear, it might be because the participants’ level or motivation was lower than those who answered positively. At the intermediate level, the time given in the FTFsa was found adequate by 81% while far fewer participants found it enough to answer the questions in the CASA. The number of intermediate level test takers who found the visuals in the FTFsa and the CASA supportive were close to each other, yet the number of them were considerably lower than the number of the pre-intermediate level participants who agreed that the visuals were beneficial. As for the responses to open-ended

questions, surprisingly, the comments on visuals were all related to the CASA at both levels. Participants 8, 16, and 23 at the pre-intermediate level stated that they liked the visual materials illustrated on the screen. The intermediate level participants made more comments on the effect of the visual materials integrated into the questions. Participants 32, 57 and 61 stated that the existence of the visual materials was what they liked best. Participant 66 said that the visuals facilitated her ability to think. Finally, participant 55 stated that having visuals was nice because it made the test more enjoyable. In short, they liked having visuals because it facilitated their task to perceive a question.

Asking whether they could easily organize their ideas during the speaking test, item 26 revealed that few students were able to do so in both test modes when examined in general. Yet, mean scores on the level-based analysis showed that participants at both levels were able to think more clearly during the FTFsa. Among numerous possible reasons, finding the FTFsa environment more relaxing, the existence of an interlocutor more being helpful, or getting anxious in an unfamiliar context (the CASA) can be considered prominent ones. It is important to consider the responses to the open-ended questions where a number of test takers reported having difficulty in organizing their thoughts in the FTFsa, though. For instance, a large number of students at the pre-intermediate level stated that they were bothered by the fact that they were extremely anxious and they could not organize their thoughts or remember and use the appropriate vocabulary to continue the conversation effectively. The lines below are examples of such comments:

> Participant 6: I was really bothered by the fact that I made mistakes I normally would not do because of my unnecessary tension. In addition, getting confused

> about the words meanings of which I did not know for sure caused a similar feeling.
>
> Participant 9: I was very nervous and tense, so I could not find words to continue my speech.
>
> Participant 11: It was bothersome that I couldn't organize my thoughts during the test.
>
> Participant 22: I was irritated since I couldn't transmit what I thought in English.

Likewise, the intermediate level test takers were discontent with the fact that they had trouble in remembering the structures or words in the FTFsa, and failed to express themselves as they would have liked to. Participant 58 wrote she was anxious because she thought in Turkish but could not transfer it to English. Two other participants, 59 and 61, stated that the words simply did not come to their mind. The excerpts below also show the high levels of anxiety which resulted in displeasure:

> Participant 32: I was tense because what I would do in the test would turn into test scores later. I wanted to demonstrate my full proficiency and this anxiety felt so bad.
>
> Participant 37: There is a certain amount of time and you have to speak within that time. I even forget what I can say since this induces anxiety.
>
> Participant 41: The test done like a one-to-one private lesson was what disturbed me the most. The speaking test should resemble the daily speech more or perhaps they can be conducted in a more convenient environment.
>
> Participant 43: I was unable to express what I wanted to say easily. I could have built more accurate sentences. I couldn't remember the words. It was nerve wracking.

The lines indicate that no matter how hard the interlocutors tried to calm the test takers down, some of them were still extremely nervous and anxious and their efforts did not actually change the reality that it was a test in the end. These findings suggest that a number of participants at both levels were unable to remember things

due to the anxiety they experienced because of various factors, which probably means that they were unable to show their full ability in speaking.

Similarly, some test takers reported having difficulty in organizing their thoughts in the CASA. Surprisingly, there were less complaints regarding this issue in the CASA than in the FTFsa and the complaints came from the intermediate level only. Participant 38 stated that she felt anxious because she could not think of anything to talk about. Participants 40 and 48 noted that they had difficulty in organizing their ideas and putting the words together due to the high level of anxiety they experienced. Likewise, participant 43 stated that she could not remember the answers.

Finally, item 27 focused on how the participants perceived the fairness of the tests. While the majority believed that the FTFsa was a fair speaking test, around 30% thought the CASA was also fair and a similar number of participants thought that it was not. As the level-based analysis revealed, except for the intermediate level test takers' answers regarding the CASA, the speaking tests were found to be fair, which means that the pre-intermediate CASA was also perceived as fair as opposed to the overall results.

Qualitative Analysis of the Open-Ended Questions in the Perceptions Questionnaires

In this section, the responses of the pre-intermediate and intermediate level participants to the open-ended questions, which are different from the actual questionnaire items in the FTFsa and the CASA perceptions questionnaires, will be reported, qualitatively analyzed and interpreted respectively. As the descriptive

statistics revealed, there are differences in test takers' perceptions of the FTFsa and the CASA. A closer look at what the test takers thought about the two modes will give a better insight into how they felt about both test modes as the information presented in this section was not included within the set of 29 questionnaire items.

To start with the pre-intermediate level test takers' perceptions of the FTFsa, it was seen that not all of the students appreciated all aspects of the FTFsa despite the fact that descriptive statistics revealed that FTFsa was found satisfactory in many cases. Below are the pre-intermediate test takers' responses showing their dissatisfaction with the test mode and interpretations of them.

> Participant 3: I was most irritated because I had to wait for my turn to come outside of the class. This caused me more stress.
>
> Participant 10: We were asked to wait for too long before our speaking turn came. It could have been organized better.
>
> Participant 11: It bothered me not to able to take the test on time and to wait.

As is clear from the excerpts above, the pre-intermediate test takers felt more anxious and irritated as they had to wait for a long time since they were invited to the classroom one by one to take the test. For each participant, the test lasted for around 6-8 minutes, and there were 17 students in one class, and 19 in the other. So it took more than half an hour for some of the students to have their turn, which possibly made them frustrated, tense or bored. Likewise, at the intermediate level, a vast number of participants found it very wearisome to spend time waiting for their turn to come. For instance, participant 45 noted that it really irritated and tired him to wait for his turn. Participant 53 said he waited for two hours before he could take the test.

Depending on the results from the quantitative analysis, it is possible to say that the test takers at both proficiency levels had less positive feelings towards the

CASA. Below is a detailed qualitative analysis of the participants' responses to the open-ended questions in the CASA perceptions questionnaires, which were intended to give a deeper insight into their feelings and opinions about this computer assisted speaking test.

To start with the positive attitudes, some of the participants seem to have liked the CASA just because they enjoyed a different experience. For instance, one participant at the pre-intermediate level said he was glad that such an interactive application was used, and another said that he enjoyed the test because it was the first test he had ever taken on computer. Similarly, participants 54 and 51 at the intermediate level noted that they enjoyed speaking actively in the test, with one of them adding that he would like to have similar tests more often to practice his speaking. Moreover, the fact that the recordings were clear and comprehensible, and the time given to answer the questions was adequate appealed to participants 44 and 47.

Despite the fact that the students had positive feelings towards the CASA, they also criticized it for a number of reasons. One of the biggest deficiencies was the technical problems, as also noted by participant 24. The following lines from the pre-intermediate level confirm the fact that the technical equipment should have been better to be able to apply such a test more efficiently:

> Participant 2: I was disturbed by other people's voices…
>
> Participant 5: We do not have the necessary technological infrastructure for such a test…
>
> Participant 10: The problems in the internet connection diverted my attention away.

> Participant 18: Since the test was done on the internet the buffering of the video clips was delayed and resulted in a waste of time. I was also very disturbed by the fact that some computers malfunctioned.

Similar to their pre-intermediate peers, the intermediate level participants made complaints about the technical problems they encountered. Although the researcher instantly responded to the technical failures during the test, it was not enough to relieve the students. The following lines reveal how disturbed they were due to these problems:

> Participant 52: I was worried due to factors such as the low speed of the internet, the fact that the pages we clicked on opened very late, and hearing other test takers' voices.
>
> Participant 53: It was difficult to get used to the system of the test. Once you clicked on the wrong place, the page would close…The computers ought to be renewed, … and the system should be simplified.
>
> Participant 62: The computers were inefficient. So, I couldn't demonstrate my full performance. The computers should be renewed and there should be a better system.

As is obvious from the test, the test takers were discontented with the technical equipment, which may be the reason why their scores on the CASA perceptions scale were relatively low. However, considering that a lot of students got high scores also in the CASA, it is possible to suggest that this did not affect all test takers' performance negatively.

Another issue that caused discomfort might be the unfamiliarity of the test mode. Some participants reported being anxious because it was the first time they had taken a test in computerized mode. It is important to note that such comments came only from the participants at intermediate level. The lines below are exemplary:

> Participant 58: I felt the anxiety of taking such a test for the first time.

> Participant 54: I was nervous since it was the first time I took this kind of test. It would be better if we had done some exercises in the previous days.

The last group of problems reported by more pre-intermediate than intermediate level participants related to being unable to answer some questions either because of their difficulty or lack of time. Some test takers complained about failing to answer some of the questions because they were difficult, the students had low listening proficiency, or there was no chance for clarification requests. A total of five pre-intermediate and two intermediate test takers noted that the time given was not enough to answer the questions although more time was given in the CASA than in the FTFsa for the same questions.

To summarize, the results of the qualitative analysis done with the open ended questions revealed that, at both levels, there were students favoring either of the test modes due to many different reasons.

The Relationship Between Speaking Anxiety and Speaking Test Anxiety, Test Mode-Related Perceptions And Test Scores

The test takers were given a questionnaire with two subsections to gather information about their speaking anxiety in general and speaking test anxiety in particular. The reliability analysis computed for the Speaking Test and Speaking Anxiety Questionnaire (see Appendices K and L) showed the Cronbach's Alpha Coefficient to be .919 for the speaking test anxiety subscale, and .944 for the speaking anxiety subscale. The information gathered from the questionnaire was compared with the information from the perceptions scales and the test scores.

Table 18 shows the relationship between speaking anxiety and speaking test anxiety, and the CASA and FTFsa perceptions; and the relationship between speaking anxiety and speaking test anxiety and test scores in the two different modes.

Table 18

The Relationships Between Speaking/Speaking Test Anxiety and Perceptions and Test Scores in Two Different Modes

Level		N	Speaking Test Anxiety		Speaking Anxiety	
			ô	Sig.	ô	Sig.
Overall	FTFsa score	64	-.14	.10	-.09	.28
	CASA score	64	-.12	.16	-.05	.50
	FTFsa Perceptions	64	-.25**	.004	-.26**	.002
	CASA Perceptions	64	-.20*	.02	-.25**	.003
Pre-intermediate	FTFsa score	26	-.01	.92	-.04	.77
	CASA score	26	.05	.70	.22	.12
	FTFsa Perceptions	26	-.09	.49	-.18	.19
	CASA Perceptions	26	-.25	.07	-.30*	.03
Intermediate	FTFsa score	38	-.21	.06	-.16	.15
	CASA score	38	-.22	,057	-.22*	.04
	FTFsa Perceptions	38	-.36**	.001	-.40**	.00
	CASA Perceptions	38	-.13	.23	-.25*	.02

Note. N= number of participants; ô = Kendall's tau correlation coefficient; *. Correlation is significant at the 0.05 level (two tailed); **. Correlation is significant at the 0.01 level (two tailed).

On the whole, both speaking test anxiety and the speaking anxiety are negatively correlated with the perceptions of both the CASA (ô= -.20, *p*(two tailed)= .02; ô=-.25, *p*(two tailed)= .003) and the FTFsa (ô= -.25, *p*(two tailed)= .004; ô=-.26, *p*(two tailed)= .002); neither type of anxiety correlates with the scores on either test mode. This shows that test takers with higher speaking test anxiety or higher

speaking anxiety felt less positively towards the FTFsa and the CASA. The fact that no significant correlation was found between speaking test anxiety or speaking anxiety and the scores indicates that there is no relationship between the scores gained and the anxiety levels of the test takers.

At the pre-intermediate level, speaking test anxiety and the perceptions of either test were found not to be significantly correlated. This shows that high speaking test anxiety is not related to the attitudes towards the test modes. Although no significant correlation was found between speaking anxiety and the FTFsa perceptions, there was a significant negative correlation (ô = -.3, *p* (two-tailed) =.034) between speaking anxiety and the CASA perceptions which means that the pre-intermediate level test takers with higher speaking anxiety tend to feel less positively towards the CASA, yet speaking anxiety and the FTFsa perceptions are not related. No significant correlation was found between either type of anxiety and the test scores, so it is clear that there is no relationship between the scores gained and the anxiety levels of the test takers at the pre-intermediate level.

As for the intermediate level, speaking test anxiety and the FTFsa perceptions were negatively correlated (ô = -.365, *p* (two-tailed) = .001), but no correlation was found between speaking test anxiety and the CASA perceptions. This finding indicates that, despite the fact that test takers' level of speaking test anxiety was moderately related to their perceptions of the FTFsa, it was not related to their perceptions of the CASA. There was a significant negative correlation between speaking anxiety and the perceptions of both test modes. So the test takers with higher speaking anxiety felt less positively towards both the CASA and the FTFsa. When the scores and the types of anxiety were compared, the only significant or

marginally significant correlations were found between the two types of anxiety and the CASA scores (see Table 18). It means that FTFsa scores are not related to anxiety levels, unlike the CASA scores at intermediate level.

The Relationship Between Computer Attitudes, Test-Mode-Related Perceptions And Test Scores

The participants completed the Computer Attitudes Questionnaire, which was aimed at collecting information about their perceived self efficacy in using computers and their attitudes towards computers. The reliability analysis run for the Computer Attitudes Questionnaire (see Appendices M and N) showed the Cronbach's Alpha Coefficient to have a value of .865.

Table 19 shows the results of comparisons between the results from the Computer Attitudes Questionnaire and test mode related perceptions questionnaires in addition to the scores from both tests modes.

As seen in the table, no significant correlation was found between the computer attitudes or the FTFsa or the CASA perceptions of the test takers on the whole and at the intermediate level. This means that test takers' attitudes towards computers are not related to their perceptions of the FTFsa or the CASA, in other words, a test taker with positive attitudes towards computers does not necessarily feel positively towards either type of test.

Table 19

The Relationship between Computer Attitudes, Perceptions and Test Scores

Level			FTFsa score	CASA score	FTFsa Perceptions	CASA Perceptions
Overall	Computer Attitudes	ô	.02	.09	-.08	-.12
		Sig.	.74	.28	.34	.13
		N	64	64	64	64
Pre-int.	Computer Attitudes	ô	.01	.24	-.19	-.26
		Sig.	.94	.08*	.16	.06*
		N	26	26	26	26
Int.	Computer Attitudes	ô	.03	.006	-.06	-.08
		Sig.	.77	.96	.58	.43
		N	38	38	38	38

Note. N= number of participants; ô = Kendall's tau correlation coefficient; Sig.= (two tailed); *= Correlation is marginally significant at the 0.05 level (two tailed).

Similarly, the FTFsa or the CASA scores and the test takers' attitudes towards computers were not significantly correlated, which indicates that a test taker who feels positively towards computers would not necessarily obtain high scores from the CASA or a test taker feeling negatively towards computers would not necessarily get a lower score than those who feel positively towards computers.

At the pre-intermediate level, there was no significant correlation between the computer attitudes and the FTFsa perceptions of the test takers. Likewise, no significant correlation was found between computer attitudes and FTFsa scores, whereas there was a marginally significant correlation between the CASA scores and the computer attitudes. However, there was a marginally significant correlation (ô = -.26, p= .060) between the CASA perceptions and the computer attitudes. This suggests that there may be a negative relationship between the way the test takers perceive the CASA and how they feel about using computers, though not a strong one.

Conclusion

In this chapter, the quantitative and qualitative data analysis results based on the data gathered from the CASAs and the FTFsas as well as the perceptions and anxiety questionnaires were reported. The analyses yielded important results which will be discussed in the next chapter along with the implications, limitations of the study, and suggestions for further research.

CHAPTER V: CONCLUSION

Introduction

This study aimed to investigate the pros and cons of face-to-face and computer assisted speaking assessment with reference to pre-intermediate and intermediate students' perceptions of these test modes, the scores they obtained, the anxiety levels, and the computer attitudes of the students in each mode. The data were collected through speaking tests and questionnaires on perceptions, speaking test and speaking anxiety, and computer attitudes. The participants were 66 EFL language learners studying at pre-intermediate and intermediate levels at Uludağ University School of Foreign Languages. Four language instructors also took part in the study as interlocutors and raters during the face-to-face and after the computer assisted tests.

The study had a counter-balanced design so there were two pre-intermediate and two intermediate groups who took either the FTFsa or the CASA first, and the remaining test second. After the first tests were conducted, the participants were administered either the FTFsa or the CASA perceptions questionnaires depending on the test they took. The process was repeated after a one-month period for the second speaking tests. Within the same time period, the participants were given two other questionnaires, namely, the speaking test and speaking anxiety questionnaire, and the computer attitudes questionnaire. Quantitative data were entered in SPSS and analyzed via descriptive and inferential statistics, and qualitative data were typed, grouped, and interpreted qualitatively.

In this chapter, the research findings will be discussed and evaluated in the light of the research questions and the relevant literature. Within the scope of the chapter, pedagogical implications, limitations of the study and suggestions for further research will also be presented.

Findings and Discussion

Performance, Reliability and Validity: The Scores Obtained in the CASA and the FTFsa

After participants at both pre-intermediate and intermediate levels took both the CASA and FTFsa, their scores were calculated and the scores they obtained in the two different test modes were compared first altogether and then separately at two levels. Before focusing on the findings, it is important to note that the inter-rater reliability scores were quite low, especially for the CASA as opposed to the findings of Jeong (2003). The CASA and the FTFsa were rated by four instructors – two at each proficiency level- which means that, for each level, the same two raters scored each test. Although the content, style and number of the questions were also identical in both test modes, the pairs of raters were found to rank the test takers inconsistently and this was more obvious in the CASA rankings, in contrast to the findings in the literature (Cheng, 2008; Kuo & Jiang, 1997; Stansfield & Kenyon, 1988). For instance, a student who was ranked the third in the class by one rater, was the 19th in the other raters' scoring sheet. The inter-rater reliability scores for the FTFsa were more promising, yet they too were not high enough to say that the test performances were evaluated as they should have been. The reason for this is possibly the absence of experience in giving and scoring speaking tests. Although all of the raters were experienced EFL instructors, none of them had previous experience with speaking

tests. It is important to recall that the speaking tests conducted within the scope of the present study were the first and only speaking tests administered ever at Uludağ University. Therefore, even though the raters were instructed on how to evaluate the speaking performances and were given clearly written guidelines, it was not enough to standardize the way they perceive and rate speech samples. As for the differences in the inter-rater reliability scores between the CASA and the FTFsa, the reason why the reliability was so much lower in the CASA than that in the FTFsa is open to debate since the questions in both test modes were identical, in addition to the fact that the raters were not allowed to negotiate about the scores in the FTFsa. The absence of visual clues might be one reason for this inconsistency, while being allowed to decide on the time and the location to assess the recordings might be another as the raters might not have paid equal or adequate attention to the task of scoring the CASA. Obviously, the language instructors would need more training in scoring the speech samples from the semi-direct speaking tests than the face-to-face interviews. The low inter-rater reliability scores also invite us to interpret the other findings in this study cautiously.

The findings revealed that there were either no correlations or only a weak correlation between scores gained on the two tests both in general and at separate levels. This result is in line with findings of Jeong (2003), who investigated whether there was a significant difference in results between a computerized oral test and the conventional face-to-face format for the sake of questioning the reliability of the computerized test. The researcher reported that there was only a weak relationship, with a correlation value of .3 between the scores from the two oral proficiency tests. Moreover, Silvester (2000) found that the ratings of applicants were different in

situations where the oral proficiency of the test takers were assessed via face-to-face or via telephonic interviews. The results of this study should be interpreted with care in the present context as it is a telephone-based interview, not a computer based speaking test. However, it is not completely irrelevant because participants' lack of practice and nonverbal cues in the telephone-based interviews also resemble the computer assisted speaking tests.

In contrast to the low correlations of scores reported in the above studies, Stansfield and Kenyon (1992) found in a study comparing the Simulated Oral Proficiency interview (SOPI), a semi-direct speaking test, and the Oral Proficiency Interview (OPI) that there was a high correlation between the scores obtained in different test modes. Similarly, a number of other studies (Kenyon & Malabonga, 2001; Qian, 2009; Shohamy, 1994; Wigglesworth and O'Loughlin, 1993) have found the concurrent validity of the two test modes to be high. In another study carried out by Xiong, Chen, Liu, and Huang (as cited in Cheng, 2008), the students were reported to have demonstrated their actual oral language proficiency through a semi-direct test, depending on the fact that a high correlation was observed between students' ranking in class and their scores which were obtained through a tape-mediated semi-direct test. The results from Surface (2009) also confirmed these results. Once again, though, it is important to keep in mind that the results of these studies may not directly relate to the present study because the studies cited above were tape-based semi-direct tests except for Kenyon and Malabonga (2001) and Qian (2009), which were computer assisted.

Although a number of studies found high correlations between scores, according to what O'Loughin (2001) plausibly suggests, getting similar scores on

two different test modes may not mean that the two kinds of test measure the same kind of ability or that the ability is measured equally, which indeed means that one of them might be lacking construct validity. In other words, the language elicited through computers may be different from that elicited via a face-to-face test and the two tests could be testing distinct components of speaking proficiency. Similarly, Chapelle and Douglas (2006) assert that performance on a computer-delivered test may fall short of reflecting the same ability as what other forms of assessment would measure if presenting items on the computer screen changes the mental processes to respond correctly to them. More specifically, Shohamy (1994) reports based on her qualitative analyses that there were differences between the communicative strategies and discourse features used in direct and semi-direct assessment of oral proficiency. O'Loughlin further points out that the spoken interaction of two people is jointly constructed so it is basically different from communicating with a machine. Therefore, researchers (Clark, 1979; O'Loughlin, 2001) caution against using the semi-direct and face-to-face oral proficiency tests interchangeably. Obviously, there might be numerous reasons to declare semi-direct speaking tests to be a form of assessment that lacks reliability and concurrent validity, as there is a possibility that the results from them would not match those from their face-to-face counterparts in terms of content, even if the scores correlate. In this study, the scores did not correlate significantly, which further emphasizes the point that the CASA cannot be a substitute for the FTFsa, at least in its current form.

However, this does not mean that we should totally get rid of the computerized speaking tests as it is still possible to get some valid information from

them related to the oral performance of the test takers. The approach Norris (2001) takes is worth considering. Norris states that:

> Language test developers should start their deliberations about speaking assessment not by asking what computers are capable of doing, but rather by asking what kinds of interpretations actually should be made about L2 speaking abilities; what kinds of evidence a test will need to provide in order to adequately inform those interpretations; and what kinds of simulation tasks will provide the required evidence. (p.103)

The choice of test mode should basically depend on what the stakeholders, i.e. teachers, hope to find out about students' performance, in other words, the specific testing needs of their institutions, as Jiang and Kuo suggests (1997).

In the present study, the aim of using computer assisted semi-direct oral assessment was to observe students' improvement over time. The speaking tests were to be progress achievement tests intended to measure the progress the students made (Hughes, 2000); in other words, whether the students had learnt certain aspects of spoken language introduced in speaking classes and could use them actively. Thus, instead of testing their overall proficiency by looking at each and every detail of spoken language, the raters actually looked for certain patterns in students' speech. No aspects other than those they had been instructed on were assessed during the tests. That is, the tests measured what they were supposed to measure. This means that both types of tests had content validity because their content constituted a representative sample of the language skills, structures and so on with which they were meant to be concerned (Hughes, 2000). Among the criteria Brown (2004, p.27) lists for a test to have face validity are a well constructed, expected format with familiar tasks, a test doable within the allotted time limit, clear items and directions,

tasks that relate to the course content, and reasonable challenge. It is also clear from the test takers' answers to item 5 of the FTFsa and the CASA perceptions questionnaires, where 34% of the participants agreed and 36% partly agreed that the CASA effectively tested what was taught in the speaking classes, and from items 11 and 24, where a considerably high number of participants were satisfied with the instructions and the time limit that the tests also had face validity.

Both the FTFsa and the CASA can be said to have construct validity as it was possible to observe the expected constructs in the responses of some students, especially of those who got higher scores and probably studied what was taught in the class in detail. Nevertheless, it is crucial to note that most of the tasks in the curriculum were monologic tasks. The computer assisted tests of oral proficiency have been shown to be valid tools for proficiency in monologic tasks in terms of the accuracy and complexity of the speech (Zhou, 2008), though not yet in conversational ones. Hence, even though the validity of a number of tasks in the CASA which were supposed to be conversational in nature is questionable, the monologic tasks (i.e. describing a picture, and giving a short presentation) which constituted most of the tasks in both test modes, seem to be appropriate devices to be used in the CASA type of tests.

For a deep insight into the distribution of the test scores, the interaction of test type, proficiency level, and the order the tests were taken were investigated. The results revealed that there was no main effect of test mode, level, or group alone, which means that none of these factors changed the average score by themselves. Nevertheless, there was an interaction between level and test mode. The pre-intermediate students were found to do better on the FTFsa than on the CASA,

whereas for the intermediate students it was vice versa. This information confirmed that the two levels were affected differently by test mode and suggested that the CASA had an impact on the pre-intermediate test takers resulting in lower performance quality.

Another result of the interaction analysis showed that test mode and group, namely, the groups who took the tests in different orders, also interacted significantly. On the whole, the groups which took the CASA first were found to do better on the FTFsa, and the groups which took the FTFsa first did better on the CASA. Obviously, both groups did better in their second test than in their first test no matter which type of test they took first, which shows that there was a *practice effect* in general. It might suggest that some students performing poorly in the CASA probably did so because they had never taken a speaking test before, or at least it was the first time they were taking a speaking test at the institution they were studying during the investigation and it was in the CASA mode. Further examination of the interaction between test mode, proficiency level, and group showed that the practice effect was only seen in the pre-intermediate level.

CASA and FTFsa Perceptions of the Test Takers

A descriptive analysis of the results from the CASA and the FTFsa perceptions questionnaires revealed that there was some divergence in perceptions of the two test modes. Adding to the body of research, the test takers were found to prefer the FTFsa over the CASA and have more positive feelings towards the former in general. However, it is important to note that it was the pre-intermediate level students who seemed to favor the FTFsa considerably more whereas the attitudes of the test takers towards both modes were not as harshly different at the intermediate

level. This might be a result of less anxiety and more self confidence that led higher proficiency test takers to perceive both modes relatively similarly. Irrespective of the proficiency level, a large amount of research conducted with different participants, tasks, methodologies, and technologies (Jeong, 2003; Luoma, 1997; Qian, 2009; Silvester, 2000; Stansfield et al., 1988, 1990; Thompson, 2007; Yu & Lowe, 2009) has presented results in support of the view that the face-to-face format is preferred over the computerized, or other semi-direct modes of oral proficiency assessment. As is clear from the descriptive analysis of the perceptions questionnaires, among the prominent reasons for this preference are the "unnatural" structure of the semi-direct speaking tests as opposed to the interactive interviews bearing a communicative nature; the presence of a live on-site interviewer that relieves the test takers in the face-to-face mode; the lack of gestures to facilitate the conversation in the semi-direct mode, and the lack of experience with the semi-direct mode, as the literature also suggests.

When the test-mode-related anxiety subscales of the two perceptions questionnaires were examined, the test takers were found be more anxious in the CASA than in the FTFsa at both levels. This is possibly because - in addition to all the aforementioned reasons that lead to a preference for the face-to-face mode - they had to deal with many things such as using the computer, familiarizing themselves with the system and trying to demonstrate their oral English proficiency at the same time. It is noteworthy that there were a number of technical problems during the CASA and the students did not have a chance to practice with the software system except for a short demonstration before the actual test due to time restrictions, which might have resulted in higher levels of anxiety in this test mode.

In line with the findings above, the test takers were found to feel more anxious before, during and after the CASA than the FTFsa in general. One reason for that might be the lack of an opportunity for clarification, repetition or restatement requests from the interviewers, which can be considered a disadvantage of the semi-direct tests (Kenyon & Malone, 2010) because the chance to ask for clarification from an on-site interviewer might help the test takers feel that they can control and correct their own performance (Silvester, 2000), while another factor could be the positive, friendly and relieving attitudes of the interlocutors in the FTFsa, as the responses to the open-ended questions at both levels revealed. Although the responses also indicated that some test takers were extremely anxious in the FTFsa as well, no matter how hard the interlocutors tried to calm them down, the test takers who received no interlocutor support in the CASA and were thus anxious outnumbered them. Interestingly, the level-based analysis showed a difference in the periods when the students at different levels were anxious. A lot of the pre-intermediate students were not anxious before the CASA while most were anxious during the CASA, but the intermediate students were anxious both before and during the computer assisted test. This might be due to the way their speaking teachers speculated about the difficulty of the test before it was administered. On the other hand, students at both levels were still anxious after the CASA though this was not the case for the FTFsa. This is possibly the result of trying something novel.

Moreover, a number of participants (33%), especially those at the pre-intermediate level (46.4%), noted that the presence of someone listening to them instead of talking to a computer relieved them. Speaking to a computer was one of the prominent causes of the discomfort almost half of the participants at both levels

felt. In addition, more than half of the test takers reported being afraid of making mistakes in the CASA, though much fewer test takers were scared by this in the FTFsa. The reason for this may be the interlocutor interference in the FTFsa, given that the interlocutors typically try to relieve test takers during the interviews. In support of this view, the literature suggests that that the examinees' satisfaction with the interviewer predicted their reactions to the face-to-face test mode (Thompson et al., 2007), and a sincere and caring manner was the most effective (Madsen, 1983). Weir (2005) suggests starting a speaking test with personal or social questions designed to decrease anxiety, similar to Luoma's (2004) proposal that there should be a warm-up section in the beginning. Also, Oya et al. (2004) recommends that the interlocutors should facilitate a more convenient testing environment to reduce the negative impact of anxiety. In line with this body of research, it is possible to say that the interviewers behaved appropriately, which resulted in less anxiety in the FTFsa in turn. Nevertheless, in the CASA perceptions questionnaire, the lack of a live interviewer was reported to be a problem which caused discomfort because the test takers had difficulty in speaking due to lack of interaction and an intimate environment. On the other hand, some participants at both levels felt more relaxed when there was no one listening to them, that is, while talking to a computer. In line with this, there were participants who were even disturbed by the presence of an interlocutor and a few others feeling constrained by the interference of the interviewers. This might be due to the differences in their personalities or learning styles. For instance, some students might be shy whereas others are extrovert, or some could be intrapersonal learners who can understand and work well by themselves whereas others are interpersonal who understand and work well with

others (Gardner, 2006, p.18), which can be another factor the test givers should keep in mind while designing tests and selecting the mode of the tests.

Another point that relates to the anxiety the students felt in the two test modes is the perceived difficulty of the tests. At both levels, the majority of the test takers thought that the CASA was more difficult, and they reported that they expected low scores from the CASA. Although many students also thought they would get low scores from the FTFsa, their number was much lower. This is surprising since both tests were identical in terms of content, style and the number of questions. Therefore, the anxiety felt or the difficulty found cannot be related specifically to the questions. Moreover, it is important to remember that their perceptions hardly correlate with their scores, which means that there is only a weak relationship between their perceptions of their performances and their genuine performances in the speaking tests. One possibility is that presenting items on the computer screen might change the mental processes to respond to them (Chapelle & Douglas, 2006), as mentioned before, so that the test takers might feel that they have to put more effort into answering a question in the computerized mode.

The participants also stated that they found it difficult to organize their ideas in the speaking tests so they were unable remember the words and structures they already knew, or transfer what they think in Turkish into English. Even though they had similar problems in both test modes, the CASA was again found more dissatisfactory in this respect. Finding the FTFsa environment more relaxing, the existence of an interlocutor being more helpful or getting anxious in an unfamiliar context (the CASA) can be considered prominent reasons for this. However, it is crucial to remember that it was just a perception question and does not reflect on

their actual performance. So, it is possible that the students were able to organize their ideas well and easily in both of the test modes. The only way to see what actually happened is to analyze the content and organization of students' responses to the test questions, which is not in the scope of this study.

The responses to the open ended questions revealed that the technical problems, which really hindered the flow of the test, were found irritating by a number of the students at both levels. Indeed, this may be the main reason why students disliked the computerized test and got more anxious during it. Despite the considerable amount of effort spent in preparing the test, there were problems with the computers because they were old, and the Internet connection was rather slow, and it was not possible to change it as it was a state university where even making a small change required a lot of effort and time. The website used for recording and storing the voices gave errors randomly at different times, causing anxiety in test takers, which was an unexpected problem beyond the researcher's means, so it is recommended that schools develop their own websites to record and store the responses to avoid such problems. Probably, conducting the speaking test with better technical equipment would yield better results in favor of the CASA, both in terms of perceptions and performance. However, it is also noteworthy that there were test takers obtaining high scores in the CASA, which might mean that the technical problems did not hinder at least some test takers' performance. As reported by a few participants, taking a computerized test for the first time could also have resulted in low performance. In other words, even if there were no technical problems, it is possible that utilizing the system would still pose a challenge and decrease the scores

gained in the CASA. The only way to overcome this is to expose students to computerized tests repeatedly.

Considering that identical questions were asked in both the CASA and the FTFsa, one would expect the perceptions regarding the questions in the two tests to be similar. However, the majority favored the FTFsa over the CASA when they were asked about the ability of a test to reflect their ability and to show one's strengths and weaknesses in speaking English, and the comprehensiveness of the test as well as the variety of the questions. Surprisingly, though, this preference was more obvious at the pre-intermediate level, because the intermediate level test takers thought neither of the tests were good tools for assessing their speaking ability and they were dissatisfied with the comprehensiveness and the diversity of both the CASA and the FTFsa. They might have expected tests which assessed their general speaking ability, but as they were told beforehand, these were achievement tests which intended to test a limited amount of knowledge the students were supposed to have gained during their speaking classes, not general proficiency tests.

Although time given to answer the test questions was found adequate by a large number of students as the descriptive statistics showed, a few test takers at both levels responding to the open ended questions thought that the time given in the CASA was not enough to answer the questions while no one reported such a problem for the FTFsa. Nevertheless, it is important to note that less time was given for the same questions in the FTFsa; it took only 6-8 minutes for each student to answer these questions. In the CASA, the participants were given half an hour in total: they first did a trial with the system, which took about ten minutes, and the remaining 20 minutes were devoted to answering the actual questions. The rationale behind giving

a lot more time in the CASA was the possibility that the internet connection or other technical deficiencies would cause students to lose time. Although students were supposed to have been provided with adequate time, some of them found it inadequate, as mentioned above. This might have resulted from the fact that students were unacquainted with the software used to administer the CASA.

Both in the CASA and the FTFsa the same visual materials, i.e., pictures, drawings, or graphics, were used with the aim of facilitating the test takers' comprehension. They were found beneficial in the FTFsa by the participants at both levels. However, only the intermediate test takers found the visuals in the CASA useful, which may have resulted from the fact that the pre-intermediate level students who found the visual materials unsupportive might have failed to answer the questions due to their lower proficiency level and believed that the visuals did not help at all. At intermediate level, it was stated that visuals facilitated the ability to think, which shows that the multiple modality of input in the CASA probably helped the tests takers demonstrate their actual proficiency. It was also stated that having visuals was nice because it made the test more enjoyable. Briefly, they liked having visuals because it facilitated their task to perceive a question via both auditory and visual modes.

In general, the majority found the CASA neither real life like nor fair whereas it was the opposite for the FTFsa, supporting Luoma (1997). This is surprising because in the body of relevant literature (Galaczi, 2010; Larson, 2000; McNamara 2008), one of the most stressed attributes of the computerized tests is their fairness. Interestingly, though, a considerably high number of pre-intermediate test takers

found the CASA to be fair. The reason why intermediate level test takers thought the CASA was unfair as opposed to their pre-intermediate peers is open to investigation.

When the participants were asked if the existence of a speaking assessment would affect their attendance rates, it was found that only the FTFsa would increase the pre-intermediate level students' attendance to the speaking classes whereas the intermediate level test takers thought that neither of the test modes would change their attendance rates. It might mean that only the FTFsa would have some positive washback effect on lower level students' participation in the classes. This might simply be happening because higher level test takers feel more confident in speaking English and may believe that they do not need extra instruction.

Up to this point, the findings showing that the test takers mostly favored the FTFsa over the CASA have been discussed. Indeed, one of the clearly striking findings is that 75% of the pre-intermediate and 47% of the intermediate level test takers preferred the FTFsa according to the results from item 22, confirming the relevant literature (Kenyon & Malabonga, 2001; Thompson et al., 2007). Nevertheless, the FTFsa had also some disadvantages in the test takers' view.

The test takers were disturbed by a number of factors in the FTFsa. The high level of anxiety of some students which emerged before the FTFsa was said to increase their nervousness, which resulted in difficulty in remembering words or structures, or organizing their speech in turn, especially as stated by intermediate level test takers. The type of anxiety reported here probably stems from a general speaking anxiety or speaking test anxiety, as some respondents complained that they were unable to relax since they knew it was a test. In this case, one would assume

that the participants would have also felt the same way before the CASA but most of them did not. Stansfield and Kenyon (1988) claimed that semi-direct tests seemed 'unnatural' and unfamiliar to the participants, which resulted in greater perceived difficulty and more nervousness than was felt in the face-to-face format. In the present study, perhaps the feeling of trying something new, that is, taking a speaking test on computer, and the fear created by this challenge was so high that the test takers did not even care about their speaking anxiety. This might have even been caused by the computer anxiety the students might have experienced.

The participants also reported having to wait for a long time, which actually meant spending more than two hours for some of them before it was their turn in the face-to-face speaking test. They noted that this increased their already existing anxiety and they started to feel rather tense after a while. Although some of the participants recommended testing half of the students another day or in another classroom with only one instructor, it does not seem to be an applicable solution to do either as both would have adverse effects on reliability of the speaking tests because of differences between raters or the times the test would be given.

Evidently, not pre-intermediate but intermediate level test takers were disturbed by the fact that one of the two interlocutors conducting the FTFsa kept taking notes as the interview went on. The students probably felt that the interlocutor was uninterested in their speech either because she found it unintelligible or just because she did not care. This might have been done unconsciously, or the interlocutor might have thought that it was necessary to grade the student without missing any details of his performance by waiting for him to leave. Since some test takers were evidently bothered by this fact, it would be better to control such

behaviors during the face-to-face speaking tests to avoid a failure in terms of reliability.

Speaking test and speaking anxiety and their relationship with the perceptions of the test modes and test scores

In order to eliminate possible factors other than test-mode specific perceptions that might affect how students feel about a test due to their general fears or tendencies, the test takers were given a questionnaire with two subsections: speaking anxiety and speaking test anxiety.

The analyses conducted to explore the relationship between speaking anxiety and speaking test anxiety, and the test-mode specific perceptions revealed that there was a significant negative correlation between speaking anxiety and the CASA perceptions at the pre-intermediate level. No significant correlation was found between speaking anxiety and the FTFsa perceptions or the test scores the pre-intermediate examinees obtained in either test mode. These findings indicate that the pre-intermediate level test takers with higher speaking anxiety tend to feel less positively towards the CASA, yet speaking anxiety and the FTFsa perceptions are not related. In line with Phillips (2005), there is no relationship between speaking anxiety and speaking test scores at this proficiency level in the present study, either, which contradicts the findings of Woodrow (2006), Park et al. (2005), Phillips (1992), and Oya et al. (2004), who found that test scores were affected by debilitating speaking anxiety, and Aydın (2006), who found that test takers who obtained high scores in overall achievement tests felt more confident and less anxious. Nor did speaking test anxiety correlate significantly with any of the

abovementioned variables. This shows that high speaking test anxiety is not related to the attitudes towards the test modes and the test scores at pre-intermediate level.

At intermediate level, a significant negative correlation was found between speaking anxiety and CASA and FTFsa perceptions. Hence, it can be stated that the test takers with higher speaking anxiety felt less positively towards both the CASA and the FTFsa. Moreover, there was a significant moderate negative correlation between speaking test anxiety and FTFsa perceptions at intermediate level but it was not related to their perceptions of the CASA. Comparing the test scores and the anxiety types, a significant negative correlation was found between the scores obtained in the intermediate CASA and the two types of anxiety, which is in line with the findings of relevant studies (Woodrow, 2006; Park et al, 2005; Aydın, 2006; Phillips 1992; Oya et al., 2004). It means that FTFsa scores were not related to the anxiety levels, unlike the CASA scores at intermediate level.

The fact that speaking anxiety and speaking test anxiety were found to be related to FTFsa perceptions at the intermediate level but not at the pre-intermediate level might suggest that the interlocutors at the pre-intermediate level were more friendly and helpful than those at the intermediate level, which might have resulted in more negative feelings towards the FTFsa and a negative correlation between speaking anxiety and FTFsa perceptions at the intermediate level, unlike the pre-intermediate level. As for the negative correlation between the CASA scores and the anxiety levels at both levels, the intermediate level participants might have failed to demonstrate their actual proficiency and lost points as they got anxious, whereas the pre-intermediate level participants, proficiency levels of whom are already low,

might have performed similarly in the CASA no matter whether they were anxious or not.

Computer attitudes and their relationship with the perceptions of the test modes and test scores

With the same aim as that of the speaking test and speaking anxiety questionnaire, the participants were asked to complete a computer attitudes questionnaire.

According to the results of their study, Taylor, Kirsch and Eignor (1999) and Powers (1999) claimed that there was no adverse relationship between computer familiarity and computer-based TOEFL test performance due to absence of experience with computers. In support of this, no relationship was found between intermediate level test takers' computer attitudes and their test scores. Neither their computer attitudes were found to be related to their test-mode specific perceptions.

The analyses showed a marginally significant correlation only between pre-intermediate level test takers' computer attitudes and their CASA perceptions (a negative correlation) in addition to their CASA scores (a positive correlation). This means that there may be a relationship between how the lower level examinees perceive the CASA and how they feel about using computers, though not a strong one. Similarly, their computer attitudes and CASA scores seem to be weakly related. Since a high score in the computer attitudes questionnaire meant that the test taker had negative attitudes towards the computer and a high computer anxiety, whereas a high score in a test is an indicator of good performance, the moderate positive correlation found between the CASA scores and computer attitudes is interesting. It

actually means that test takers with negative attitudes towards computers performed better in the Computer Assisted Speaking Assessment, which might mean that negative attitudes towards computers which might have resulted in computer anxiety acted as facilitating anxiety and helped the test takers focus on the tasks better. The reason that computer attitudes were found to relate to test scores only in one proficiency level is open to investigation. It may have resulted from differences in computer familiarity and anxiety levels at two different levels. If this is the case, it would be better to investigate the variations in computer familiarity of the populations for whom the computerized tests are intended, as Chapelle and Douglas (2006) propose.

Pedagogical Implications

To start with a difficulty language teachers might face when they choose to use a computerized speaking test, it is worth clarifying that setting up a software system that would be used school-wide requires a noticeable amount of time and effort, though not a big economical power. Setting up the CASA system, as a language teacher with little technical knowledge, I had difficulty in managing some processes. First, one of the two basic software components in this study, Moodle, required an appropriate webserver which is compatible with the structure of Moodle. After a lot of individual effort which resulted in failure, I was able to find a server, install Moodle, and integrate VoiceThread into Moodle with the help of a friend. It is important to note that even two friends, who were graduates of computer related departments at universities, failed to complete this process with success until the third person mentioned above could achieve it. Second, due to the lack of a professional technical department, I had to deal with the technical problems at the

computer laboratory at Uludağ University personally because the some computers needed to be changed, repaired, modified, or supported with extra materials as most of them had viruses, software and internet connection problems, and there were no microphones or headphones at all. The technical problems continued even after everything was completed and while the students were taking the speaking tests. For instance, some computers started to malfunction all of a sudden, the quality of the internet connectivity descreased uncontrollably, or the VoiceThread website started not to respond on one computer while it was still working on other computers. The problems experienced during the tests irritated the test takers despite the efforts to overcome them instantaneously. Problems like the ones above show that language teachers should think twice ensure that they can receive adequate technical support before ever starting to use computers to assess speaking as it is not as simple as it sounds.

On the other hand, in the light of the findings of the study, it is possible to conclude that creating local computerized tests of speaking is not beyond the capability of the language instructors, yet, the stakeholders should be careful while converting the conventional face-to-face speaking tests into computer assisted assessment of speaking, as the two might give very different rankings to the students. The administrators and instructors should carefully decide on the parts of the speaking curriculum they want to evaluate, decide whether these attributes can be assessed via computers, and give a chance to the instructors and the students to try the software systems used. It should also be ensured that the technical equipment, such as the headphones, microphones, computers and the internet, are working properly, since technical problems seem to relate to the anxiety students felt during

the CASA. Other anxiety raising factors, such as hearing other people's voices and having difficulty in getting used to the system should also be eliminated to allow students show their full performance, as anxiety can have a debilitating effect for some students. To make the CASA type of tests more appealing for students, it might be a good idea to insert video clips in addition to the pictures or graphics used in the study so that the students may feel that a conversation is being simulated, which might compensate somewhat for the lack of interactiveness inherent to the semi-direct oral assessment. The test takers could also be required to interact with each other via the internet using video-conferencing tools, which would totally eliminate the most prominent disadvantage of a semi-direct test of speaking.

In order avoid unreliable rankings of students resulting from the low inter-rater reliability in test modes in this study, the schools must pay attention that the instructors who would evaluate students' speaking performances are not biased in any way or inexperienced in conducting and rating speaking tests. Comprehensive in-service training is necessary to avoid this problem, especially in Turkey because few institutions assess or even give instruction on speaking ability in the grammar-based EFL setting of the country, and those who do so pay little attention to the way speaking is evaluated.

In addition to their advantages for the test givers, such as being economical and time-saving, computer assisted speaking tests utilized as achievement tests would also help the students get used to semi-direct forms of assessment, which they will encounter later in their life. Moreover, the recently-popular language portfolios can be supported by language learners' voice recordings, which would provide more authentic and convincing evidence about their speaking proficiency than bare scoring

charts do. Despite its current disadvantages, the computerized speaking tests can be used instead of the face-to-face tests on condition that the abovementioned deficiencies of the computerized speaking tests are eliminated. Beyond doubt, these problems would be solved with ease with the rapid technological developments in a short period of time.

The results of the study also suggest that there is a clear preference for the FTFsa over the CASA, which is assumed to be a result of unfamiliarity with the latter. Nevertheless, students' needs and interests should also be taken into consideration when possible. That is, when feasible and economical, face-to-face tests should be the first choice. Administering face-to-face speaking tests has also its drawbacks, though. As reported by the participants, having to wait for a long period of time until their turn come results in discomfort among test takers. Moreover, it was seen that test takers tend to learn the content and the style of the speaking test from their peers who take the test before them, which would decrease the reliability of the test. Test takers can be put into separate classrooms, or waiting rooms, while waiting for their turn. The use of mobile communication devices should also be prevented and test takers should be accompanied by a responsible instructor. To economize on time spent on testing, the tests can be conducted on different days. Obviously, this process would require extra human resources and available classrooms. Conducting speaking tests with the help of one single instructor is also an option, but not a favorable one, due to reliability concerns.

The findings also revealed that test takers disliked some interlocutor behaviours, such as taking notes during the interview. This suggests that the interlocutors should be strictly trained on how to behave during the interviews. To

prevent negative effects of the anxiety raising nature of the face-to-face tests, the tests can be conducted at different times during the class hours without informing the students that it is a test and the performances can be rated using a detailed rubric; however, it would possibly be an unfair way of assessment as the students would be tested with different questions.

The low inter-rater scores in this study show that language teachers might give very different rankings to test takers, both in the CASA and in the FTFsa, even though they are provided with a detailed list of answers and band descriptors. Therefore, an in-depth rater traning, which can be done using real speech samples and comparisons between the scores raters give, is also indispensably necessary even if language teachers insist on conducting face-to-face oral assessment.

Limitations of the study

There are several limitations of this study which suggest that the findings should be interpreted cautiously. As mentioned before, one of the major weaknesses of the study is the raters' lack of experience and the rather low inter-rater reliability. As a result of this, the actual success of the test takers was possibly probed inaccurately and the analyses done using these scores might have resulted in incorrect interpretations.

The lack of clarity of the rubric and the band descriptors used in the study and the fact that the raters did not have time to practice using them in training sessions are the other limitations that might have resulted in low inter-rater reliability.

Technical problems experienced during the CASA pose another thread to the validity of the interpretations of the scores as well as the perceptions and anxiety

questionnaires. If schools or researchers intend to use a computer assisted test, they should initially ensure that the technical infrastructure would not cause any difficulty.

Another limitation of the study is the monologic nature of the most of the questions in the speaking tests. Though this is an inherent characteristic of semi-direct oral assessment, it is not plausible to say that such a test can perfectly assess every aspect of the test takers' speaking ability.

A further limitation is the setting where the study was conducted. Only one school and four classes were included in the study and the number of participants was quite low – 13 students in each pre-intermediate group and 19 students in each intermediate group with 66 student participants in total - to be able make nation-wide generalizations. It is possible for such a study to give different results in different settings due to the interests, capabilities, and beliefs of the participants.

Another limitation is the lack of control of the students as they waited for their turn to take to FTFsa. It was observed that students waiting for their turn interacted with their peers who already took the test to get information about the flow of the test. Although the test takers had been informed about the content of the speaking tests and they had the necessary sources beforehand, i.e. speaking textbooks, their performances might have been influenced by the extra information they received from their peers.

Time constraints were also among the limitations of the study because the preparation and piloting of the instruments, the preparation of eight different speaking tests and the training materials for the raters were completed in a short

period of time. The raters should have been trained better by helping them practice rating speech samples and standardizing their rating techniques; however it was impossible due to the time limitations. Within this time period, the needs of the students might have been neglected, as well. For instance, the only chance they found to try the CASA system was just before they took the test. Ideally, the test takers should have been given a few trials to get used to it before the actual test.

Suggestions for further research

Based on the findings and the limitations of the study, some suggestions can be made for further research. This study was one of the few instances which have looked at the use of computerized speaking tests as progress achievement tests. It is true that dealing with technology is challenging, but the instructors ought to get used to it as it is one of the indispensible parts of the near future. Therefore, other studies investigating how computer assisted speaking assessment can be effectively integrated into the local curriculum would be highly valuable.

The study can be replicated with better equipment, experienced raters, more participants, or at different settings to find whether the two types of speaking tests have similar effects and results at different conditions.

Third, a communicative version of semi direct speaking tests where test takers interact with and see each other, i.e. utilizing video conferencing or virtual worlds, can be developed and evaluated using a similar methodology to this study or in more detail. This would reduce the disadvantages of lack of interaction inherent to the taped or computerized speaking tests.

Fourth, exploring which question types elicit more and better information about the test takers' progress in terms of oral competency can also be another way of getting a better insight into the nature of the semi-direct speaking tests and to validate them.

Moreover, since there is the possibility that different cognitive processes involved in different test modes might be affecting the speech produced, investigating how the test takers' cognitive processes differ in the two modes, if they do at all, might lead to new research on the types of test items and modes that would eliminate any negative effects of computerized speaking assessment.

Finally, how the speech samples from a semi-direct speaking test should be evaluated, in other words, what should be or can be expected from such tests and where they can be appropriately used can be investigated.

Conclusion

This study revealed that the face-to-face and computer assisted tests of speaking might give very different rankings to students, particularly if the raters and the students are unfamiliar with the latter. It was also revealed that the test takers at both levels clearly preferred the face-to-face mode of speaking assessment over a computerized version for various reasons. Also, the test takers at different proficiency levels were found to value different aspects of the test modes and there were some students who favored the CASA, though not numerous. Nevertheless, their perceptions were not found to determine success in the speaking tests. The speaking test anxiety and speaking anxiety questionnaires showed that students' perceptions of the test modes, especially at intermediate level, were related to their

speaking or speaking test anxiety. Moreover, there was a relationship between these anxiety types of the intermediate level test takers' and their computer assisted speaking assessment scores. Finally, pre-intermediate level test takers' CASA perceptions and CASA scores were found to be related to their computer attitudes, though in the opposite direction.

Most importantly, this study has highlighted the importance of the quality of the tools to be used at oral assessment as well as the processes oral assessment involves. The study had also drawn attention to language teachers' training and experience in testing speaking, the quality of the technical equipment used in computerized testing, the clarity of the scoring rubric, and the students' experience with the test technique as important aspects of speaking assessment. Insights from this study are hoped to prove useful in designing new speaking assessment tools.

REFERENCES

Aydın, S., Yavuz, F. & Yesilyurt, S. (2006). Test anxiety in foreign language learning. *Balikesir University, Journal of Social Sciences Institute, 9*(16), 145 – 160.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, UK: Oxford University Press.

Bailey, K. M. (2006). Issues in teaching speaking skills to adult ESOL learners. In J. Comins, B. Garner & C. Smith (Eds.), *Review of Adult Learning and Literacy, Volume 6: Connecting Research, Policy, and Practice* (pp. 113-164). Mahwah, NJ: Lawrence Erlbaum Associates.

Bernstein, J., De Jong, J. H. A. L., Pisoni, D., & Townshend, B. (2000). Two experiments on automatic scoring of spoken language proficiency. In P. Delcloque (Ed.), *Proceedings of InSTIL2000: Integrating Speech Technology in Learning*, *University of Abertay Dundee, Scotland,* 57-61.

Birjandi, P., & Alemi, M. (2010). The impact of test anxiety on test performance among Iranian EFL learners. *Broad Research in Artificial Intelligence and Neuroscience*, *1*(4), 44-58.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.

Bygate, M. (2001). Speaking. In R.Carter & D.Nunan (Eds.), *The Cambridge guide to teaching English to speakers of other languages* (pp.14-20). Cambridge: Cambridge University Press.

Chaney, A. L., & Burk, T. L. (1998). *Teaching oral communication in grades K-8*. Boston: Allyn & Bacon.

Chang, Y., J., Wu, C.,T., & Ku, H. Y. (2005). The Introduction of electronic portfolios to teach and assess English as a foreign language in Taiwan. *TechTrends, 49*(1), 30-35.

Chapelle, C. A., & Douglas, D. (2006). *Assessing language through computer technology*. Cambridge, UK: Cambridge University Press.

Clariana, R., & Wallace, P. (2002). Paper-based versus computer-based assessment: Key factors associated with the test mode effect. *British Journal of Educational Technology, 33*(5), 593-602.

Clark, J. L. D. (1979). Direct versus semi-direct tests of speaking proficiency. In Briere, E. J. & Hinofotis, F. B (Eds.), *Concepts in language testing: Some recent studies* (pp. 35–49). Washington, DC: Teachers of English to Speakers of Other Languages..

Clark, J. L. D. (1986). Development of a tape-mediated ACTFL/ILR scale-based test of Chinese speaking proficiency. In C. W. Stansfield (Ed.), *Technology and language testing* (pp. 129-146). Washington, DC: Teachers of English to Speakers of Other Languages.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (2002). *Dictionary of language testing*. Cambridge: Cambridge University Press.

Fall, T., Adair-Hauck, B. & Glisan, E. (2007). Assessing students' oral proficiency: A case for online testing. *Foreign Language Annals*, *40*(3), 377-406.

Fulcher, G. (2003). *Testing Second Language Speaking*. London: Longman/Pearson Education.

Fujii, Y. (1993). Construction of a test influence inventory (TII). *Japanese Journal of Psychology, 65*(2), 135-139.

Galaczi, E. D. (2010). Face-to-face and computer-based assessment of speaking: Challenges and opportunities. In L. Araújo (Ed.), *Computer-based assessment (CBA) of foreign language speaking skills* (pp. 29-52). Luxembourg: Publications Office of the European Union.

Gardner, H. (2006). *Multiple intelligences: New horizons*. New York: Basic Books.

Goldberg, A., L., & Pedulla, J., J. (2002). Performance differences according to test mode and computer familiarity on a practice Graduate Record Exam. *Educational and Psychological Measurement, 62*(6), 1053-1067.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal, 70*(2), 125-132.

Huang, H. T. D., & Hung, S. T. A., (2010). Implementing electronic speaking portfolios: perceptions of EFL students. *British Journal of Educational Technology, 41*(5), 84-88.

Hughes, A. (2000). *Testing for language teachers*. Cambridge: Cambridge University Press.

Jeong, T. (2003). *Assessing and interpreting students' English oral proficiency using d-VOCI in an EFL context*. Unpublished doctoral dissertation. Ohio State University, Columbus.

Kay, R. H. (1993). An exploration of theoretical and practical foundations for assessi ng attitudes toward computers: The computer attitude measure (CAM). *Computers in Human Behavior, 9,* 371-386.

Kenyon, D., M., & Malabonga,V. (2001). Comparing examinee attitudes toward computer assisted and other oral proficiency assessments. *Language Learning & Technology, 5*(2), 60-83.

Kenyon, D., & Malone, M. (2010). Investigating examinee autonomy in a computerized test of oral proficiency. In L. Araújo, (Ed.), *Computer-based assessment (CBA) of foreign language speaking skills* (pp. 1-28). Luxembourg: Publications Office of the European Union.

Kirsch, I., Jamieson, J., Taylor, C., & Eignor, D. (1998). *Computer familiarity among TOEFL examinees.* (TOEFL Research Report 59; ETS Research Report 98-6). Princeton, NJ: Educational Testing Service.

Kuo, J., & Jiang, X., (1997). Assessing the Assessments: The OPI and the SOPI. *Foreign Language Annals, 30*(4), 503-512.

Larson, J. W. (2000). Testing oral language skills via the computer. *CALICO Journal, 18*(1), 53-66.

Lazaraton, A. (1996). Interlocutor support in oral proficiency interviews: The case of CASE. *Language Testing, 13*(2), 151-172.

Loyd, B.H., & Gressard, C. (1984). Reliability and factorial validity of Computer Attitude Scales. *Educational and Psychological Measurement, 44*, 501-505.

Luoma, S. (1997). *Comparability of a tape-mediated and a face-to-face test of speaking: A triangulation study.* Unpublished master's thesis, Centre for Applied Language Studies, University of Jyvaskyla, Jyvaskyla, Finland.

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

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

McNamara, T., Hill, K., & May, L. (2002). Discourse and assessment. *Annual Review of Applied Linguistics, 22*, 221-242.

McNamara, T. (2008). *Language testing*. Oxford: Oxford University Press.

Norris, J. M. (2001). Concerns with computerized adaptive oral proficiency assessment. *Language Learning & Technology, 5*(2), 99-105.

O'Loughlin, K. J. (1997). *Direct and semi-direct tests of spoken language*. Unpublished doctoral dissertation, University of Melbourne, Melbourne, Australia.

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

Oya,T., Manalo, E., & Greenwood, J. (2004). The influence of the personality and anxiety on the oral performance of Japanese speakers of English. *Applied Cognitive Psychology, 18*, 841-855.

Park, H., & Lee, R.A. (2005). L2 Learners' anxiety, self-confidence and oral performance. *Proceedings of the 10th Pan-Pacific Association of Applied Linguistics (PAAL), Japan*, 197-208. Retrieved June 01, 2011 from http://www.paaljapan.org/resources/proceedings/PAAL2010/pdfs/hyesook.pdf

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

Phillips, L. (2005). *A study of the impact of foreign language anxiety on tertiary students' oral performance*. Unpublished master's thesis, University of Hong Kong, Hong Kong.

Powers, D. E. (1999). *Test anxiety and test performance: Comparing paper-based and computer-adaptive versions of the GRE general test.* (Research Rep. No. 99-15). Princeton, NJ: Educational Testing Service.

Qian, D. D. (2007). Assessing university students: Searching for an English language exit test. *RELC Journal*, *38*(1), 18–37.

Qian, D. D. (2009). Comparing direct and semi-direct modes for speaking assessment: Affective effects on test takers. *Language Assessment Quarterly. 6*(2), 113-125.

Shohamy, E. (1994). The validity of direct versus semi-direct oral tests. *Language Testing, 11*, 99-123.

Silvester, J., Anderson N., Haddleton, E., Cunningham-Snell, N., & Gibb, A. (2000). A cross model comparison of telephone and face- to-face selection interviews in graduate recruitment. *International Journal of Selection and Asseesment, 8*(1), 16-21.

Speaking criteria. (n.d.). Retrieved March 5, 2011, from http://www.pearsonlongman.com/languageleader/pdf/teacher-help/speaking_criteria.pdf

Stansfield, C. W. (1990). An evaluation of simulated oral proficiency interviews as measures of oral proficiency. In J. E. Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics: Linguistics, language teaching and language acquisition: The interdependence of theory, practice and research*. Washington DC: Georgetown University Press, 1990.

Stansfield, C.W., & Kenyon, D. M. (1988). *Development of the Portuguese speaking test*. Washington, DC: Center for Applied Linguistics.

Stansfield, C.W., Kenyon, D.M., Paiva, R., Doyle, F., Ulsh, I., & Cowles, M. A. (1990). The development and the validation of the Portuguese Speaking Test. *Hispania, 73*(3), 641-651.

Stansfield, C.W., & Kenyon, D. M. (1992). The development and validation of a simulated oral proficiency interview. *The Modern Language Journal, 76*(2), 129-141.

Stansfield, C. W., & Kenyon, D. (1993). Development and validation of the Hausa speaking test with the ACTFL proficiency guidelines. *Issues in Applied Linguistics, 4*(1), 5-31.

Stricker, L. J., & Attali, Y. (2010). *Test takers' attitudes about the TOEFL IBT™.* (TOEFL iBT™ Report No. iBT-13). Princeton, NJ: Educational Testing Service.

Surface, E. A., Harman, R. P., Watson, A. M., & Thompson, L.F. (2009, April). *Are human and computer-administered interviews comparable?* Paper presented at the 24th annual meeting of the Society for Industrial and Organizational Psychology, New Orleans, LA.

Taylor, C., Kirsch, I., & Eignor, D. (1999). Examining the relationship between computer familiarity and performance on computer-based language tasks. *Language Learning, 49*(2), 219-274.

Thompson, L. F., Surface, E. A., & Whelan, T. J. (2007, April). *Examinees' reactions to computer-based versus telephonic oral proficiency interviews.* Paper presented at the 22nd annual meeting of the Society for Industrial and Organizational Psychology, New York, NY.

Thornbury, S. (2005). *How to teach speaking*. Harlow, England: Pearson-Longman.

Usó-Juan, E., Martínez-Flor, A., & Alcón-Soler, E. (2006). Approaches to language learning and teaching: Towards acquiring communicative competence through speaking. In E. Usó Juan & A. Martínez Flor (Eds.), *Current trends in the development and teaching of the four language skills* (pp. 139-157*)*. Berlin: Mouton de Gruyter.

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

Wigglesworth, G., & O'Loughlin, K. (1993). An investigation into the comparability of direct and semi-direct versions of an oral interaction test in English. *Melbourne Papers in Language Testing*, *2*(1), 61-71.

Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal, 37*(3), 308-328.

Xi, X. (2008). *Investigating the criterion-related validity of the TOEFL® speaking scores for ITA screening and setting standards for ITAs.* (TOEFL iBT™ Report No. iBT-03). Princeton, NJ: Educational Testing Service.

Yu, X., & Lowe, J. (2009). Computer assisted testing of spoken English: A study to evaluate the SFLEP college English oral test system in China. *Systemics, Cybernetics, and Informatics, 7*(3), 33-38.

Zhou, Y. (2008). A comparison of speech samples of monologic tasks in speaking tests between computer-delivered and face-to-face modes. *JLTA Journal, 11*, 189-208.

APPENDICES

APPENDIX A: Intermediate FTFsa 2 Questions

0. WARM-UP (not more than 30-40 seconds)

Please introduce yourself. (Name, age, hobbies etc.)

1. TRAVEL AGENT

Imagine that you are working at a travel agency.

1. You offer some VACATION PACKAGES to your customers. SCAN the brochures QUICKLY to see them.
2. Listen to the person speaking. She is looking for an IDEAL VACATION. Choose a SUITABLE VACATION for her.
3. **Tell her the DETAILS and why she should prefer it. Describe the FEATURES OF PLACES, make RECOMMENDATIONS, use STRONG ADJECTIVES just like you learnt in the speaking class.**

SCRIPTS (for the interlocutor to read)

a) I'm a business woman and work really hard during the year. I have a **one-week** holiday in **July**. I want to go somewhere I can both relax and have fun, probably **Turkey**. I have been wondering Turkish cities for a long time. I would like to spend my days **at sea** and **historical places.** For nights, I'd prefer **night clubs**. And I have a limited budget: **only 900** euros. What do you suggest? (Aegean Region)

b) I'm a **retired** person so I want somewhere **quiet**. I am interested in **historical places**. **Camping or small boutique hotels** are fine for me. The big **5 star hotels** are usually too crowded and noisy, I don't want to spend my time with a lot of people around. Also, it would be perfect if there is a place with **thermal baths**. When you get older they are really good for you, you know! So what do you think, Is there a suitable tour for me? (Black Sea Region)

c) Isn't Turkey the home of Ottoman Empire? I would really like to spend a few days visiting the most famous **historical places,** you know, **museums, palaces, mosques**, the places **where the great war**s were done… It would be nice to try the **Turkish bath**, too. I'm sure I would like to buy a lot Turkish stuff, so I will need to go to **bazaars** as well. Unfortunately, I have **a 7 day** holiday only. Can I do all these in such a short time? (Marmara Region)

d) I am looking forward to having a vacation. I'm so tired! I want to go somewhere I can **relax**. The only activities I want to do are **fishing** lazily and **swimming**. I heard that the **thermal baths** are a good way of relaxing too. After I rest a little, I would like to do some

climbing, I love mountains! Finally, I don't like hot places. Can you offer me a place where I can do all these? (East Anatolian Region)

e) Hi! I won a holiday check from a TV show. It's worth **2000 euros**. I want to spend it somewhere sunny and energetic. I love night life so the place should definitely have **night clubs**. I have 11 days and I want to do as many things as possible. I would like to have a **cruise**, see natural beauties like **waterfalls, caves, lakes**… I want to do some sports too. For example, **diving, trekking, or rafting**… And please make sure that the hotel is a very very good one. Where can I go? (Mediterranean Region)

East Anatolia Region

Cities: Ağrı, Iğdır, Erzurum, Van

Accommodation: four-star hotels

Activities/ places to go: Thermal baths, trekking, religious historical monuments (e.g.Oltu Church), palaces (e.g. İshak Pasha), climbing, fishing, swimming

Duration: 5 days

Best time: From May to November

Price: €800 (euro)

South East Anatolia Region

Cities: Gazi Antep, Şanlıurfa, Adıyaman, Mardin

Accommodation: 4 star hotels

Activities/ places to go: Sightseeing, museum (archeology museum), Lake (Balıklıgöl), Nemrut mountain (historical statues and nice sunset, Kasımiye Madrassa (Kasımiye Medresesi), lots of traditional food options

Duration: 4 days

Best time: From May, September, November

Price: € 600

Mediterranean Region

Cities: Antalya, Adana, Burdur, Kahramanmaraş

Accommodation: boutique and beach resort hotels

Activities/ places to go: diving, swimming, trekking and bird observation, Turkish cuisine (traditional food), rafting , cruise tour, visiting caves (Damlataş, İnsuyu), waterfalls (e.g. Düden), castles (there are eight castles in Kahramanmaraş), mosques, bridges, Lakes (Eğirdir, Kovada, Burdur, Salda), night clubs

Duration: 10 days

Best time: From May to September

Price: € 1600

Aegean Region

Cities: İzmir, Aydın, Muğla, Denizli

Accommodation: 4 and 5 star hotels

Activities/ places to go: Bodrum, Dalaman, Datça, Dalyan, Marmaris for scuba diving and swimming, Ephesus antic city, antic cities in Denizli (Tripolis, Hierapolis), Pamukkale travertines, trekking and cycling, paragliding (yamaç paraşütü), Rodos island daily tour, visiting caves (e.g. İnkaya)

Duration: 6 days

Best time: From May to September

Price: € 900

Marmara Region

Cities: İstanbul, Bursa, İznik (Nicea), Çanakkale

Accommodation: 4 and 5 star hotels

Activities/ places to go: Ottoman palaces (Dolmabahçe, Topkapı), Bosporus (Boğaziçi) tour, museums, Rumeli Hisarı, islands (Gökçeada, Bozcaada), tombs (e.g. Yeşil türbe), historical mosques (Ulucami, Sultan Ahmet), Turkish bath, Troy horse, bazaars in Bursa, İstanbul and Çanakkale, Assos antic city, Nusret mine layer ship, skiing, trekking, swimming, cable car tour (teleferik), İznik china (traditional Turkish porcelain), night clubs and bars

Duration: 7 days

Best time: four seasons

Price: € 1000

Blacksea Region

Cities: Safranbolu, Rize, Trabzon, Sinop, Artvin

Accommodation: boutique hotels, tents

Activities/ places to go: Safranbolu houses, Sinop castle, Erfelek waterfalls, camping on plateaus (e.g. Ayder yaylası), original Turkish tea, rafting, thermal baths, trekking, bird observation, sightseeing, swimming

Duration: 6 days

Best time: from May to September

Price: € 600

2. TALKING ABOUT PERSONALITY, JOB, COMPANY

Imagine that you are WORKING IN A COMPANY and you are talking to a new friend about your PERSONALITY AND JOB.

First, talk about your PERSONAL QUALITIES, your strengths and weaknesses that affect your work

Secondly, tell the DEPARTMENT you are working in and DESCRIBE WHAT YOU DO there.

Finally, DESCRIBE THE COMPANY you are working in.

IMPORTANT: Speak as much as possible and use the vocabulary/ expressions you learnt in the speaking class.

3. MAKING A REQUEST AND JUSTYFYING IT

Imagine that you WORK IN AN OFFICE and you are talking to your colleagues. MAKE THREE REQUESTS and JUSTIFY YOUR REASONS. Use the ideas given below to make your requests.

Help finish writing this report

Clean my desk

Get print outs of reports

Lend me a stapler

Put the files into folders

Interpret/ explain the table for me

Need a new printer


Canon

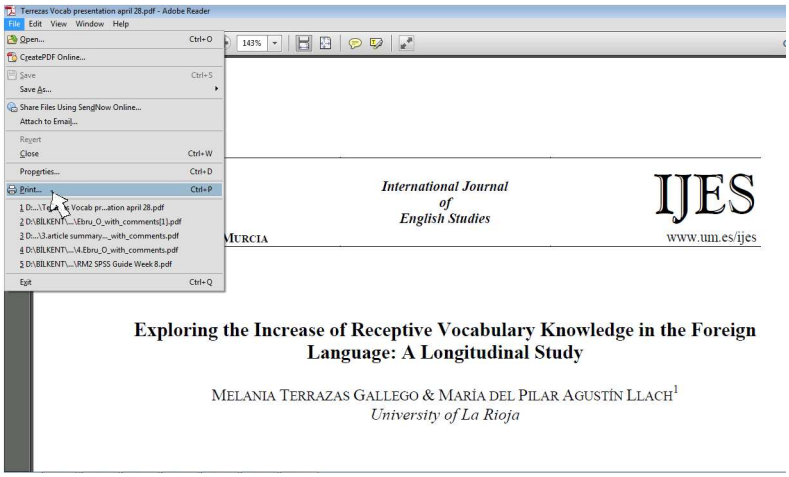

TABLE A: SCALE CONVERSION RATIOS

Desired drawing scale \ Published drawing scale	No.1 scale (1:32)	O scale (1:48)	S scale (1:64)	OO scale (1:76)	HO scale (1:87.1)	TT scale (1:120)	N scale (1:160)	Z scale (1:220)
No.1 scale (1:32)		150%	200%	238%	272%	375%	500%	688%
O scale (1:48)	67%		133%	158%	181%	250%	333%	458%
S scale (1:64)	50%	75%		119%	136%	188%	250%	344%
OO scale (1:76)	42%	63%	84%		115%	158%	211%	289%
HO scale (1:87.1)	37%	55%	73%	87%		138%	184%	253%
TT scale (1:120)	27%	40%	53%	63%	73%		133%	183%
N scale (1:160)	20%	30%	40%	48%	54%	75%		138%
Z scale (1:220)	15%	22%	29%	35%	40%	55%	73%	

APPENDIX B: Intermediate CASA 2 Questions

Page 1:

There are a few buttons you must use in this test.

1) **Comment** (The first step before "record" button)

2) **Record** (To start recording your voice, you must click on it)

3) **Stop recording** (When you finish your answer, you must click on it)

4) **Save** (DO NOT FORGET IT! To save your answer, you must click on it)

5) **Next** (To go to the next page, you must click on it. It is at the **left bottom of this webpage)**

IMPORTANT: BEFORE YOU START SPEAKING, REMEMBER TO CLICK ON THE " Sign In or Register" BUTTON AT THE BOTTOM RIGHT CORNER OF THE VIDEO.

You will have four questions in this test.

For your first question, please go to the next page.

GOOD LUCK :)

Page 2:

QUESTION 1:

Please introduce yourself briefly. (Name, age, department, hobbies, etc.)

Page 3:

QUESTION 2:

Imagine that you are working at a travel agency.

1. You offer some vacation packages to your customers. Scan the boxes quickly to see the vacation packages.
2. Listen to the person speaking. She is looking for an ideal vacation. Choose a suitable vacation for her.

3. **Tell her the details and why she should prefer it. Describe the features of places, make recommendations, use strong adjectives just as you learnt in the speaking class.**

South East Anatolia Region

Cities: Gazi Antep, Şanlıurfa, Adıyaman, Mardin

Accommodation: 4 star hotels

Activities/ places to go: Sightseeing, museum (archeology museum), Lake (Balıklıgöl), Nemrut mountain (historical statues and nice sunset, Kasımiye Madrassa (Kasımiye Medresesi), lots of traditional food options

Duration: 4 days

Best time: From May, September, November

Price: € 600

Marmara Region

Cities: İstanbul, Bursa, İznik (Nicea), Çanakkale

Accommodation: 4 and 5 star hotels

Activities/ places to go: Ottoman palaces (Dolmabahçe, Topkapı), Bosporus (Boğaziçi) tour, museums, Rumeli Hisarı, islands (Gökçeada, Bozcaada), tombs (e.g. Yeşil türbe), historical mosques (Ulucami, Sultan Ahmet), Turkish bath, Troy horse, bazaars in Bursa, İstanbul and Çanakkale, Assos antic city, Nusret mine layer ship, skiing, trekking, swimming, cable car tour (teleferik), İznik china (traditional Turkish porcelain), night clubs and bars

Duration: 7 days

Best time: four seasons

Price: € 1000

Aegean Region

Cities: İzmir, Aydın, Muğla, Denizli

Accommodation: 4 and 5 star hotels

Activities/ places to go: Bodrum, Dalaman, Datça, Dalyan, Marmaris for scuba diving and swimming, Ephesus antic city, antic cities in Denizli (Tripolis, Hierapolis), Pamukkale travertines, trekking and cycling, paragliding (yamaç paraşütü), Rodos island daily tour, visiting caves (e.g. İnkaya)

Duration: 6 days

Best time: From May to September

Price: € 900

Blacksea Region

Cities: Safranbolu, Rize, Trabzon, Sinop, Artvin

Accommodation: boutique hotels, tents

Activities/ places to go: Safranbolu houses, Sinop castle, Erfelek waterfalls, camping on plateaus (e.g. Ayder yaylası), original Turkish tea, rafting, thermal baths, trekking, bird observation, sightseeing, swimming

Duration: 6 days

Best time: from May to September

Price: € 600

Page 4:

QUESTION 3:

Imagine that you are working in a company and you are talking to a new friend about your personality and job.

First, talk about your PERSONAL QUALITIES, your strengths and weaknesses that affect your work

Secondly, tell the DEPARTMENT you are working in and DESCRIBE WHAT YOU DO there.

Finally, DESCRIBE THE COMPANY you are working in.

IMPORTANT: Speak as much as possible and use the vocabulary/ expressions you learnt in the speaking class!

Page 5:

QUESTION 4:

Imagine that you work in an office and you are talking to your colleagues. **MAKE <u>THREE</u> REQUESTS and JUSTIFY YOUR REASONS**. Use the ideas given below to make your requests.

IMPORTANT: Use the expressions you learnt in the speaking class to make requests and justify your reasons!

Help finish writing this report

Clean my desk

Lend me a stapler

Get print outs of reports

Put the files into folders

Interpret/ explain the table for me

TABLE A: SCALE CONVERSION RATIOS

Desired drawing scale	Published drawing scale: No.1 scale (1:32)	O scale (1:48)	S scale (1:64)	OO scale (1:76)	HO scale (1:87.1)	TT scale (1:120)	N scale (1:160)	Z scale (1:220)
No.1 scale (1:32)		150%	200%	238%	272%	375%	500%	688%
O scale (1:48)	67%		133%	158%	181%	250%	333%	458%
S scale (1:64)	50%	75%		119%	136%	188%	250%	344%
OO scale (1:76)	42%	63%	84%		115%	158%	211%	289%
HO scale (1:87.1)	37%	55%	73%	87%		138%	184%	253%
TT scale (1:120)	27%	40%	53%	63%	73%		133%	183%
N scale (1:160)	20%	30%	40%	48%	54%	75%		138%
Z scale (1:220)	15%	22%	29%	35%	40%	55%	73%	

CONGRATULATIONS! You have finished the test!

Click "SUBMIT ALL AND FINISH"

APPENDIX C: Rating Scale

Test type:	Face-to-face ☐	Computer-assisted ☐	The test took _____minutes.	
Name of the student:			Class:	
Name of the rater:			Date of the test:	____/ /2011

	FLUENCY	**PRONUNCIATION**	**ACCURACY**	**VOCABULARY**	**COHERENCE/ DISCOURSE**	Total grade: (out of **25**)
Points given: (out of **5**)						

APPENDIX D: Band Descriptors

Points	FLUENCY	PRONUNCIATION	ACCURACY	VOCABULARY	COHERENCE/ DISCOURSE
1	There are clear examples of communication breakdown. S cannot perform the tasks properly and communicates with much difficulty.	Pronunciation may cause miscomprehension and misunderstanding.	Grammar is not sufficient and it may take a long time to frame utterances or sometimes cannot do it at all.	Very limited vocabulary : S cannot use more than a few words or cannot use words with their correct meanings.	S has few, if any, communicative strategies.
2	S finds it difficult to complete extended utterances. S finds it difficult to perform the tasks.	Pronunciation should not be a problem but may occasionally result in misunderstanding.	S has difficulty in choosing and using grammar structures, but still can do it.	S can only use frequently used basic words accurately. S has difficulty in choosing and using advanced vocabulary.	S has limited communicative strategies.
3	S rarely hesitates, for example, especially while producing very long utterances. S can communicate throughout out the tasks.	The student may have some pronunciation problems but it should not cause communication breakdown	Grammar is adequate and any mistakes made do not result in significant breakdown of communication.	S can accurately use high frequency words as well as words presented in speaking courses. Vocabulary is adequate and any mistakes made do not result in significant breakdown of communication.	S shows some knowledge of communicative strategies and is able to form coherent utterances, shows some skill in turn-taking
4	S does not hesitate. S performs the tasks and communicates comfortably	Student's pronunciation does not impede comprehension.	Grammar and vocabulary is varied and mostly used correctly.	S can use some low frequency words as well as high frequency words accurately. Vocabulary is varied and mostly used correctly	S shows clear knowledge of communicative strategies and is able to form long and coherent utterances, shows some mastery in turn-taking.
5	S does not hesitate at all. Student's speech is perfectly smooth. S performs the tasks and communicates with ease.	S has no pronunciation errors.	S can use a variety of structures accurately.	Correct and varied vocabulary: S can use low frequency (advanced) words without making errors, can paraphrase when s/he cannot find the exact word.	S shows mastery in communicative strategies (e.g. cohesion, coherence, turn-taking.

APPENDIX E: Expected Answers for the Second Intermediate CASA and FTFsa Test Questions

1. TRAVEL AGENT

Student imagines that s/he works at a travel agency.

1. You offer some VACATION PACKAGES to your customers. SCAN the brochures QUICKLY to see them.
2. Listen to the person speaking. She is looking for an IDEAL VACATION. Choose a SUITABLE VACATION for her.

   The examinee should choose the most **appropriate** vacation pack. **If not**, after s/he describes it, you can say that it is not the vacation you want and ask for another one. At the end of the scripts you have, there is the name of vacation the examinee **must** offer.

3. **Tell her the DETAILS and why she should prefer it. Describe the FEATURES OF PLACES, make RECOMMENDATIONS, use STRONG ADJECTIVES just like you learnt in the speaking class.**

Features:	**Recommendations:**
It's a good place to go if….	You **really ought to..**
It's handy for…	You **should definitely**
It's popular for…	The place **is well worth**…
It's famous for…	You **certainly mustn't**…
You can find/see…. There	You **have to…**

Strong adjectives:
Furious, huge, starving, fantastic, terrible, fascinating, tiny, terrified, exhausted
(**NOT**! Tired, bad, interesting, angry, big, scared, hungry, good, small)

2. TALKING ABOUT PERSONALITY, JOB, COMPANY

Imagine that you are WORKING IN A COMPANY and you are talking to a new friend about your PERSONALITY AND JOB.

1. First, talk about your PERSONAL QUALITIES, your strengths and weaknesses that affect your work

Creative, reliable, methodical, flexible, well-organized, confident, determined, analytical, sociable, efficient

A good listener, good with computers, good at solving problems, can overcome challenges, can work under pressure, able to meet deadlines, good at communicating with people, a good decision maker

2. Secondly, tell the DEPARTMENT you are working in and DESCRIBE WHAT YOU DO there.

I work in the (sales department)

I'm in charge of …
My job involves…
I'm responsible for…
My main responsibility is to..
I'm interested in…
I'm mainly concerned with…

3. Finally, DESCRIBE THE COMPANY you are working in.

The company was founded/established in ..
It's based in..
The main activities of the company are..
It produces/supplies/exports/ manufactures…
It's one of the leading/ at the forefront of…
It has an annual turnover of…
It's headed by…
It's organized into three divisions.. / made up of three departments

3. MAKING A REQUEST AND JUSTYFYING IT

Student should MAKE THREE REQUESTS and JUSTIFY her /his REASONS using the ideas given below

Help finish writing this report
Clean my desk
Get print outs of reports
Lend me a stapler
Put the files into folders
Interpret/ explain the table for me
Need a new printer

Requests:

Would you mind… ing?
Could you possibly…?
Can I ask you to..?
I'd appreciate if you could…
I wonder if you could…?
I'd be grateful if you could..

For justification:

I **could really do with** a hand…

It would help a lot if…..

I'm in danger of..

I **may not finish… if I don't get help**

APPENDIX F: CASA Perceptions Questionnaire

To the attention of the participants,

Composed of two sections, this questionnaire has been prepared to gather information about the foreign language learners' attitudes towards the computer assisted speaking tests. The information obtained from the questionnaire will be used to support a research study conducted at MA TEFL program at Bilkent University. There are no correct or wrong answers in this questionnaire. Please mark the option that best reflects you and respond to all questions, this is rather important for the validity and reliability of the questionnaire. The responses you give to the questions will be kept strictly confidential.

I would like to remind that your responses are quite valuable to the study and I thank you in advance for your time and effort.

Ebru Öztekin
Graduate student
MA TEFL, Bilkent University

Informed consent form:

I understood the content and purpose of the questionnaire. I agree to complete the questionnaire and to let the researcher use my responses in the scientific study on condition that my information is kept confidential.

Participant name/ surname: _____ Signature:

Class: **:** _____

Gender/ Age **:** _____

Date **:** _____

This questionnaire is composed of three pages and two sections.
The options in the questionnaire are as below:

Strongly agree
Agree
Partly agree
Partly disagree
Disagree
Strongly disagree

Please go to the next page to start answering the questions.

FIRST SECTION:

Please mark the option that best describes you for each statement.

	Strongly agree	Agree	Partly agree	Partly disagree	Disagree	Strongly disagree
1. The speaking test was very difficult.						
2. I felt rather tense and anxious before the speaking test.						
3. I felt tense and anxious during the speaking test.						
4. I felt tense and anxious after the speaking test.						
5. The speaking test effectively tested what was taught in speaking classes or in the speaking sections of other classes.						
6. I felt very relaxed before the speaking test.						
7. I was very afraid of making mistakes during the speaking test.						
8. The speaking test will increase my attendance to speaking classes.						
9. The fact that I responded to a computer did not represent a real-life speaking experience I can have.						
10. It was irritating that I couldn't ask for clarification from the test giver.						
11. The amount of instructions given during the speaking test was too much.						

12. I could flexibly respond to the questions asked in the speaking test.						
13. I don't think that I can get a good mark from the speaking test.						
14. It relieved me to see that no one was listening to me during the speaking test.						
15. The speaking test helped me fully reflect my speaking ability.						
16. The speaking test will increase my attendance to the classes where speaking is practiced.						
17. The speaking test was a good tool for me to show my speaking ability.						
18. The fact that our speaking will be tested motivates me in terms of speaking English.						
19. The speaking test helped to decrease my fears about speaking English.						
20. The speaking test was comprehensive enough.						
21. There was no interaction during the speaking test.						
22. I would like to have my speaking tests in computerized format from now on.						
23. I think I can get a good mark from the speaking test.						
24. Adequate time was given to answer each question in the speaking test.						
25. The visual support materials helped me answer the questions.						
26. I could easily organize my thoughts in the speaking test.						
27. I think the speaking test is not a fair one.						

28. There were adequate amount and variety of questions in the speaking test to test my speaking ability.						
29. The speaking test allowed me to show my strong and weak points in speaking English.						

SECOND SECTION:

Please answer all questions shortly.

What irritated you the most in the test?

_____

_____.

What did you best like about the test?

_____

_____.

Is there anything you want to add?

_____

_____.

This is the end of the questionnaire. Thanks again for your participation ☺

Contact information for your questions and suggestions:

E-mail: educationline.tr@gmail.com

APPENDIX G: Bilgisayar Destekli Konuşma Sınavlarıyla İlgili Tutum Ölçeği

Katılımcıların dikkatine,

İki bölümden oluşan bu anket, yabancı dil öğrencilerinin bilgisayar destekli konuşma sınavlarıyla ilgili tutuları hakkında bilgi toplamak için hazırlanmıştır. Anketten elde edilen bilgiler Bilkent Üniversitesi Yabancı Dil Olarak İngilizce Öğretimi Yüksek Lisans (MA TEFL) Bölümünde yapılmakta olan bir araştırmanın içeriğine destek olması amacıyla kullanılacaktır. Bu ankette doğru ya da yanlış cevaplar yoktur. Lütfen soruları sizi en iyi yansıtan şıkkı seçerek işaretleyiniz ve tüm soruları cevaplayınız; bunu yapmanız anketin geçerliliği ve güvenilirliği açısından oldukça önemlidir. *Anket* sorularına verdiğiniz cevaplar kesinlikle *gizli tutulacaktır***.**

Vereceğiniz cevapların bu çalışma için çok değerli olduğunu hatırlatarak ayırdığınız zaman ve emeğiniz için şimdiden çok teşekkür ederim.

Ebru Öztekin
Yüksek lisans öğrencisi
MA TEFL, Bilkent Üniversitesi

Aydınlatılmış onam formu:

Anketin içeriğini ve amacını anladım. Anketi cevaplamayı ve bilgilerimin gizli tutulması şartıyla cevaplarımın ilgili bilimsel çalışmada kullanılmasını kabul ediyorum.

Katılımcının adı/ soyadı: _____ İmza:

Sınıfı/Şubesi: **:** _____

Cinsiyet/ Yaş **:** _____

Tarih **:**_____

Bu anket üç sayfadan ve iki bölümden oluşmaktadır.

Anketteki seçenekler aşağıdaki gibidir:

Kesinlikle katılıyorum
Katılıyorum
Kısmen katılıyorum
Kısmen katılmıyorum
Katılmıyorum
Kesinlikle katılmıyorum

Ankete başlamak için lütfen bir sonraki sayfaya geçiniz.

BİRİNCİ BÖLÜM:

Lütfen her bir soru için sizi en iyi yansıtan seçeneği işaretleyin.

	Kesinlikle katılıyorum	**Katılıyorum**	**Kısmen katılıyorum**	**Kısmen katılmıyorum**	**Katılmıyorum**	**Kesinlikle katılmıyorum**
1. Konuşma sınavı çok zordu.						
2. Konuşma sınavı öncesinde aşırı derecede gergin ve endişeli hissettim.						
3. Konuşma sınavı sırasında gergin ve endişeli hissettim.						
4. Konuşma sınavı sonrasında gergin ve endişeli hissettim.						
5. Konuşma sınavı konuşma dersleri ve diğer derslerin konuşma bölümlerinde öğretilen şeyleri iyi bir şekilde sınadı.						
6. Konuşma sınavı öncesinde çok rahat hissettim.						
7. Konuşma sınavı sırasında hata yapmaktan çok korktum.						
8. Konuşma sınavı benim konuşma derslerine katılımımı arttıracaktır.						
9. Konuşma sınavında bir bilgisayara cevap vermem gerçek hayatta yaşayabileceğim bir konuşma deneyimini yansıtmadı.						
10. Konuşma sınavında sınav görevlisinden soruları netleştirmesini isteyememem sinir bozucuydu.						
11. Konuşma sınavında verilen yönergeler çok fazlaydı.						

12. Konuşma sınavında sorulara esnek bir biçimde cevap verebildim.					
13. Konuşma sınavından iyi bir not alabileceğimi düşünmüyorum.					
14. Konuşma sınavında birinin beni dinlemiyor olduğunu görmek beni rahatlattı.					
15. Konuşma sınavı İngilizce konuşma becerimi tam anlamıyla yansıtmamı sağladı.					
16. Konuşma sınavı konuşma pratiği yapılan İngilizce derslerine katılımımı arttıracaktır.					
17. Konuşma sınavı İngilizce konuşma yeteneğimi göstermem için iyi bir araçtı.					
18. Konuşma sınavı yapılacak olması beni İngilizce konuşma konusunda motive ediyor.					
19. Konuşma sınavı İngilizce konuşma konusundaki korkularımın azalmasına yardımcı oldu.					
20. Konuşma sınavı yeterince kapsamlıydı.					
21. Konuşma sınavında hiç iletişim yoktu.					
22. Bundan sonra gireceğim konuşma sınavlarının bilgisayar destekli olmasını isterim.					
23. Konuşma sınavından iyi bir not alabileceğimi düşünüyorum.					
24. Konuşma sınavında her soruyu cevaplamak için yeterince zaman verildi.					
25. Konuşma sınavındaki görsel destekler cevap vermemi kolaylaştırdı.					
26. Konuşma sınavında düşüncelerimi kolaylıkla organize edebildim.					

27. Konuşma sınavının adil bir sınav olmadığını düşünüyorum.						
28. Konuşma sınavında İngilizce becerimi test etmek için yeterince sayıda ve çeşitte soru vardı.						
29. Konuşma sınavı İngilizce konuşma konusundaki güçlü ve zayıf noktalarımı göstermeme izin verdi.						

İKİNCİ BÖLÜM:

Lütfen tüm soruları kısaca cevaplayın.

Sınavda en çok neyden rahatsız oldunuz?

_____

_____.

Sınavda en çok neyi sevdiniz?

_____

_____.

Eklemek istediğiniz başka bir şey var mı?

_____

_____.

Anketiniz bitmiştir. Katılımınız ve katkınız için tekrar teşekkürler ☺

Soru ve önerileriniz için iletişim adresi:

E-posta: educationline.tr@gmail.com

APPENDIX H: FTFsa Perceptions Questionnaire

To the attention of the participants,

Composed of two sections, this questionnaire has been prepared to gather information about the foreign language learners' attitudes towards the face-to-face speaking tests. The information obtained from the questionnaire will be used to support a research study conducted at MA TEFL program at Bilkent University. There are no correct or wrong answers in this questionnaire. Please mark the option that best reflects you and respond to all questions, this is rather important for the validity and reliability of the questionnaire. The responses you give to the questions will be kept strictly confidential.

I would like to remind that your responses are quite valuable to the study and I thank you in advance for your time and effort.

Ebru Öztekin
Graduate student
MA TEFL, Bilkent University

Informed consent form:

I understood the content and purpose of the questionnaire. I agree to complete the questionnaire and to let the researcher use my responses in the scientific study on condition that my information is kept confidential.

Participant name/ surname: _____ Signature:

Class: **:** _____

Gender/ Age **:** _____

Date **:** _____

This questionnaire is composed of three pages and two sections.
The options in the questionnaire are as below:

Strongly agree
Agree
Partly agree
Partly disagree
Disagree
Strongly disagree

Please go to the next page to start answering the questions.

FIRST SECTION:

Please mark the option that best describes you for each statement.

	Strongly agree	Agree	Partly agree	Partly disagree	Disagree	Strongly disagree
1. The speaking test was very difficult.						
2. I felt rather tense and anxious before the speaking test.						
3. I felt tense and anxious during the speaking test.						
4. I felt tense and anxious after the speaking test.						
5. The speaking test effectively tested what was taught in speaking classes or in the speaking sections of other classes.						
6. I felt very relaxed before the speaking test.						
7. I was very afraid of making mistakes during the speaking test.						
8. The speaking test will increase my attendance to speaking classes.						
9. The fact that I responded to a test giver did not represent a real-life speaking experience I can have.						
10. It was irritating that I couldn't ask for clarification from the test giver.						
11. The amount of instructions given during the speaking test was too much.						
12. I could flexibly respond to the questions asked in the speaking test.						

13. I don't think that I can get a good mark from the speaking test.						
14. It relieved me to see someone listening to me during the speaking test.						
15. The speaking test helped me fully reflect my speaking ability.						
16. The speaking test will increase my attendance to the classes where speaking is practiced.						
17. The speaking test was a good tool for me to show my speaking ability.						
18. The fact that our speaking will be tested motivates me in terms of speaking English.						
19. The speaking test helped to decrease my fears about speaking English.						
20. The speaking test was comprehensive enough.						
21. There was no interaction during the speaking test.						
22. I would like to have my speaking tests in face-to-face format from now on.						
23. I think I can get a good mark from the speaking test.						
24. Adequate time was given to answer each question in the speaking test.						
25. The visual support materials helped me answer the questions.						
26. I could easily organize my thoughts in the speaking test.						
27. I think the speaking test is not a fair one.						
28. There were adequate amount and variety of questions in the speaking test to test my speaking						

ability.						
29. The speaking test allowed me to show my strong and weak points in speaking English.						

SECOND SECTION:

Please answer all questions shortly.

What irritated you the most in the test?

_____

_____.

What did you best like about the test?

_____

_____.

Is there anything you want to add?

_____

_____.

This is the end of the questionnaire. Thanks again for your participation ☺

Contact information for your questions and suggestions:

E-mail: educationline.tr@gmail.com

APPENDIX I: Yüzyüze Yapılan Konuşma Sınavlarıyla İlgili Tutum Ölçeği

Katılımcıların dikkatine,

İki bölümden oluşan bu anket, yabancı dil öğrencilerinin yüzyüze yapılan konuşma sınavlarıyla ilgili tutuları hakkında bilgi toplamak için hazırlanmıştır. Anketten elde edilen bilgiler Bilkent Üniversitesi Yabancı Dil Olarak İngilizce Öğretimi Yüksek Lisans (MA TEFL) Bölümünde yapılmakta olan bir araştırmanın içeriğine destek olması amacıyla kullanılacaktır. Bu ankette doğru ya da yanlış cevaplar yoktur. Lütfen soruları sizi en iyi yansıtan şıkkı seçerek işaretleyiniz ve tüm soruları cevaplayınız; bunu yapmanız anketin geçerliliği ve güvenilirliği açısından oldukça önemlidir. ***Anket*** sorularına verdiğiniz cevaplar kesinlikle ***gizli tutulacaktır*.**

Vereceğiniz cevapların bu çalışma için çok değerli olduğunu hatırlatarak ayırdığınız zaman ve emeğiniz için şimdiden çok teşekkür ederim.

Ebru Öztekin
Yüksek lisans öğrencisi
MA TEFL, Bilkent Üniversitesi

Aydınlatılmış onam formu:

Anketin içeriğini ve amacını anladım. Anketi cevaplamayı ve bilgilerimin gizli tutulması şartıyla cevaplarımın ilgili bilimsel çalışmada kullanılmasını kabul ediyorum.

Katılımcının adı/ soyadı: _____ İmza:

Sınıfı/Şubesi: **:** _____

Cinsiyet/ Yaş **:** _____

Tarih **:**_____

Bu anket üç sayfadan ve iki bölümden oluşmaktadır.
Anketteki seçenekler aşağıdaki gibidir:

Kesinlikle katılıyorum
Katılıyorum
Kısmen katılıyorum
Kısmen katılmıyorum
Katılmıyorum
Kesinlikle katılmıyorum

Ankete başlamak için lütfen bir sonraki sayfaya geçiniz.

BİRİNCİ BÖLÜM:

Lütfen her bir soru için sizi en iyi yansıtan seçeneği işaretleyin.

	Kesinlikle katılıyorum	Katılıyorum	Kısmen katılıyorum	Kısmen katılmıyorum	Katılmıyorum	Kesinlikle katılmıyorum
1. Konuşma sınavı çok zordu.						
2. Konuşma sınavı öncesinde aşırı derecede gergin ve endişeli hissettim.						
3. Konuşma sınavı sırasında gergin ve endişeli hissettim.						
4. Konuşma sınavı sonrasında gergin ve endişeli hissettim.						
5. Konuşma sınavı konuşma dersleri ve diğer derslerin konuşma bölümlerinde öğretilen şeyleri iyi bir şekilde sınadı.						
6. Konuşma sınavı öncesinde çok rahat hissettim.						
7. Konuşma sınavı sırasında hata yapmaktan çok korktum.						
8. Konuşma sınavı benim konuşma derslerine katılımımı arttıracaktır.						
9. Konuşma sınavında bir sınav görevlisine cevap vermem gerçek hayatta yaşayabileceğim bir konuşma deneyimini yansıtmadı.						
10. Konuşma sınavında sınav görevlisinden soruları netleştirmesini isteyememem sinir bozucuydu.						
11. Konuşma sınavında verilen yönergeler çok fazlaydı.						

12. Konuşma sınavında sorulara esnek bir biçimde cevap verebildim.					
13. Konuşma sınavından iyi bir not alabileceğimi düşünmüyorum.					
14. Konuşma sınavında birinin beni dinliyor olduğunu görmek beni rahatlattı.					
15. Konuşma sınavı İngilizce konuşma becerimi tam anlamıyla yansıtmamı sağladı.					
16. Konuşma sınavı konuşma pratiği yapılan İngilizce derslerine katılımımı arttıracaktır.					
17. Konuşma sınavı İngilizce konuşma yeteneğimi göstermem için iyi bir araçtı.					
18. Konuşma sınavı yapılacak olması beni İngilizce konuşma konusunda motive ediyor.					
19. Konuşma sınavı İngilizce konuşma konusundaki korkularımın azalmasına yardımcı oldu.					
20. Konuşma sınavı yeterince kapsamlıydı.					
21. Konuşma sınavında hiç iletişim yoktu.					
22. Bundan sonra gireceğim konuşma sınavlarının yüzyüze olmasını isterim.					
23. Konuşma sınavından iyi bir not alabileceğimi düşünüyorum.					
24. Konuşma sınavında her soruyu cevaplamak için yeterince zaman verildi.					
25. Konuşma sınavındaki görsel destekler cevap vermemi kolaylaştırdı.					
26. Konuşma sınavında düşüncelerimi kolaylıkla organize edebildim.					

27. Konuşma sınavının adil bir sınav olmadığını düşünüyorum.						
28. Konuşma sınavında İngilizce becerimi test etmek için yeterince sayıda ve çeşitte soru vardı.						
29. Konuşma sınavı İngilizce konuşma konusundaki güçlü ve zayıf noktalarımı göstermeme izin verdi.						

İKİNCİ BÖLÜM:

Lütfen tüm soruları kısaca cevaplayın.

Sınavda en çok neyden rahatsız oldunuz?

_____

_____.

Sınavda en çok neyi sevdiniz?

_____

_____.

Eklemek istediğiniz başka bir şey var mı?

_____

_____.

Anketiniz bitmiştir. Katılımınız ve katkınız için tekrar teşekkürler ☺

Soru ve önerileriniz için iletişim adresi:

E-posta: educationline.tr@gmail.com

APPENDIX J: Responses to the Open-Ended Questions in the CASA and the FTFsa Perceptions Questionnaires (Turkish)

PRE-INTERMEDIATE LEVEL

Yüzyüze Konuşma Sınavına yönelik olumlu tutumlar (Positive Attitudes Towards the FTFsa)

Rahatsız olmadım.. Resimlerini(1)

Öğretmenlerimin beni rahatlatmaya çalışması (2)

Hocaların konuştuklarımı anlıyormuş gibi davranmaları beni mutlu etti. (3)

Öğrendiğimiz bazı konuları kapsaması ve hocalarımın beni dikkatli bir şekilde dinlemesini (4)

Hocaların güler yüzlü olup beni rahatlatması (5)

Normal bir sınava göre hocaların tutumu kat kat daha iyiydi. Anlayışlı ve sabırlıydılar. (6)

Sınav sırasında herhangi bir durum rahatsız etmedi. Sınav görevlileri çok anlayışlıydı bu yüzden sınav sırasında rahat hissettim kendimi (7)

Soruların çok kolay olduğunu ve tam bizim konuşabileceğimiz seviyeydi. (8)

Hocalarımın anlayışlılığını. Güleryüzlü olmaları ve dinlendiğimi göstermeleri ayrı bir güven verdi. (9)

Yüzyüze sessiz ortamda İngilizce konuşabilmek güzeldi. (10)

Gergin bir ortam gibi durmasına karşın samimi sayılırdı. (11)

Telaffuz hataları yapılsa bile öğretmenler dediklerimi anladı. (12)

Birinin beni dinlemesi çok iyi oldu. En azından az da olsa rahatlattı. Konuştukça hoşuma gitti. (13)

Kısa sürmesini sevdim (14)

Soruların net olması güzeldi. (15)

pelin ve Şeyda hocayı. Onların güleryüzü karşısında rahat bir şekilde konuştum diye düşünüyorum. (16)

karşılıklı yüzyüze konuşuyor olmak gerçekçi bir hava kattı. (18)

Cevap veremeyeceğimden korktum ama rahatça cevaplayabildim. Yüzyüze olmak ve tanıdık hocaların olması rahatlatıcıydı (19)

Sorular günlük konuşmada karşılaşılan konuşmaları içeriyordu. Ayrıca derste üzerinde sıkça durduğumuz pratiğini yaptığımız konular soruldu. (21)

Tanıdığım ve sevdiğim hocaların sınava girmesi beni mutlu etti. (22)

Karşımda birinin beni dinliyor olmasını sevdim (24)

Hocaların çok ilgili bir şekilde beni dinlemesi beni çok mutlu etti (25)

Elimize verilen kağıtlarla aşamalı bir şekilde ilerleyebilmemiz gzeldi. Görsel olması işimi kolaylaştırdı. (26)

konu seçeneğimiz fazlaydı. Bu sayede her alanda hünerlerimizi gösterebildik. (28)

Yüzyüze Konuşma Sınavına yönelik olumsuz tutumlar (Negative Attitudes Towards the FTFsa)

Ben konusurken bir yerden not verilmesi (2)

sınavda en çok dışarıda sıranın gelmesini beklemekten rahatsız oldum. Bu benim daha çok stres yapmama neden oldu. (3)

Konuşma anında çok fazla heyecanlanmamdan ve bazı kelimelerin anlamını bilmeyişimden rahatsız oldum. (4)

Yalnızca beklemek (5)

Gereksiz heyecanımdan dolayı normalde yapmayacağım hatalar yapmam beni ciddi bir şekilde rahatsız etti. Aynı zamanda kelime anlamını tam olarak bilmediğim kelimeleri karıştımam da aynı duruma yol açtı (6)

Biraz heyecandan ve kendimi endişeli hissetmemden (8)

Çok heyecanlı ve gergindim. Bu yüzden aklıma konuşmamı sürdürecek kelimeler gelmemesi beni çok rahatsız etti. (9)

Konuşma sırası gelene kadar çok bekletildik. Daha iyi organize edilebilirdi. (10)

Sınava zamanında giremeyip beklemek beni rahatsız etti. Sınav esnasında düşüncelerimi organize edememek de rahatsız ediciydi. (11)

Özel konulardan yapılmasındansa be daha çok gündelik konular hakkında yapılmasını tercih ederdim.Öğretmenlerin konuşanlar takıldığında yardımcı olmaları gerekir bence sadece beklediler. Sohbet havasında geçmedi sınav sadece öğrenci konuştu. (12)

Sınavda en çok heyecanımdan rahatsız oldum. Konuşunca heyecana kapılıyorum ve konuşmamı engelliyor. Bunu nasıl yenebilirim bilmiyorum. Sürekli hata yapmaktan korktum. Cevap verememdiye korktum. (13)

Birinin beni dinliyor olması beni rahatsız etti (15)

Sadece sıramı beklemek biraz gerici bir durumdu (21)

Aklımda oluşmuş bazı fikirleri İngilizce olarak anlatamamaktan rahatsız oldum. (22)

Direktifler biraz fazlaydı. (25)

Tanımadığım bir denetleyicinin beni dinliyor olması beni rahatsız etti. Derste ve öğretmenlerimle konuşurken daha rahat konuşabiliyorum. (26)

Bilgisayar Destekli Konuşma Sınavına Yönelik Olumlu Tutumlar (Positive Attitudes Towards the CASA)

Rahatsız oldugum bisey yoktu Rahat bir ortam vardı (1)

Başkasına değil de bilgisayara karşı konuşmak beni daha rahat hissettirdi. (2)

soruların yazılı olarak gösterilmesini (4)

Sorular güzeldi. (6)

Ekrandaki yardımcı resimleri (8)

genel olarak sınavın kendisi güzeldi sınavı sevdim (9)

Kendi kendimi ölçtüğümü düşündüm bilgisayarla baş başa kalınca. Ancak mekanik araçlarla iletişim kurmak benim için zor oldu. Çok zevk almamama rağmen ilk için iyi bir deneyimdi. İlerdeki önemli sınavların da böyle olacağını göz önünde bulundurarak daha fazla yapılmasını umuyorum. (10)

Aslında değerlendirilmeseydim eğlenceliydi. Bilgisayara konuşmak sıkıntılı ve zor ama pratik için iyiydi. (11)

Bilgisayara karşı verdiğim ilk sınav olduğu için zevkliydi. (14)

Kimsenin beni dinlemediğini bilmek güzeldi (15)

Görsel öğeler olması ve kulaklığı sevdim (16)

Herkesin kendi sınavıyla meşgul olması cevap verirken kendimi daha rahat hissetmemi sağladı. Sınavın içeriği de güzeldi.Bu sınav sayesinde speakingde ve özellikle listeningde ne kadar yetersiz olduğumu anladım. Başarılı bir çalışma olmuş fakat ben kendi adıma sınavın bizden beklentilerini karşılayamadım. (22)

Böyle bir interaktif uygulama yapılmasını sevdim. (23)

Soruları anlamada sıkıntı çekmedim ve rahat cevap verdim (25)

Görsel öğeler olmasını sevdim. Bu sınav sayesinde speaking konusunda ne kadar yetersiz olduğumu anladım. Bu alandaki pratiğimi arttırmam gerektiğini fark ettim. (26)

Sorular günlük sorunlardan geldi. Dilin seviyesi orta düzeydi. Konuşma adına güven artıcıydı. (28)

Bilgisayar Destekli Konuşma Sınavına Yönelik Olumsuz Tutumlar (Negative Attitudes Towards the CASA)

baskalarının seslerini duymaktan rahatsız oldum (2)

İnternet bağlantısı çok yavaştı. Karşım da insan olmayınca konuşmakta sıkıntı çektim. (3)

Yeterli altyapıya sahip değiliz boyle bir sınav için. Olmasın bence bir daha (5)

Bilgisayara karşı konuşmak beni rahatsız etti. Karşımda beni dinleyen biri olmaması kötüydü.Bence yüzyüze konuşmak daha anlamlıydı. (6)

İletişimin olmamasından rahatsız oldum.bence sınavlar karşılıklı olmalı. Bilgisayar destekli olması beni rahatsız etti. (7)

Ben sınavda en çok biraz heyecanlılığımdan rahatsız oldum (8)

İnternet bağlantısında sorun olması dikkatimi dağıttı. (10)

Düşüncelerimi organize edememekten ve söyleyecek şey bulamamaktan çok rahatsız oldum. Bireysel durumlar. (11)

Internet, teknik aksaklık (12)

Kimsenin beni dinlemiyor olması. Hiçbirşeyin aklıma gelmemesi. Bence konuşma sınavları bilgisayar karşısında olmamalı. (13)

Soruları yapamamandan (14)

Soruları tam oalrak anlayamamak ve bunu soracak birinin olmaması, cevapların yanlış olduğun udüşünmek beni rahatsız etti. (15)

Süre yeterli değildi. Karşımda biri olmadığı için duygularımı gereğince ifade edemedim. (16)

Sınavda bilgisayarla konuşmak rahatsız ediciydi. Kelimeler tam anlaşılmamış olabilir. Hiçbirşeyi sevmedim. (17)

Internet üzerinden gerçekleşiyor olması videoların geç dolmasına ve zaman kaybına yol açtı. Ayrıca bir takım bilgisayar aksaklıklarının yaşanmasından oldukça rahatsız oldum. (18)

Sorular uzun cevap gerektiren sorulardı ama süre ise azdı. Karşında bilgisayar olması da hiç samimi bir ortam yaratmadı. (19)

Demek istediklerimi tam olarak ifade edemedim. (20)

Her soruya cevap verirken düşünmemiz için yeterince zaman yoktu. Herkes aynı anda konuştuğu için konsantre olamadım. Açıkçası hiçbirşeyi sevmedim. Olumlu hiçbir yanı yoktu. Bilgisayarla yapılan sınavın gerçek seviyeyi ölçtüğünü düşünmüyorum. Not verilirken bu sınav baz alınmasın. Hocalarla yüzyüze mülakat daha iyi olacaktır.(21)

Listeningim iyi olmadığı için soruları anlayamadım. Anlayamadığım bazı sorulara cevap vermek zorunda kalmam beni zorladı. Cevap veremediklerimi de boş bırakmak zorunda kaldım. (22)

İnternetin yavaş olması ve sorunun içeriğini anlayamamam. (23)

Teknik eksiklikler (24)

Sınavda gerçek bir diyalog iletişimi olmadı.(25)

Süre azdı. Karşımda biri olmayışından dolayı düşüncelerimi tam ifade edebildiğimi düşünmüyorum. (26)

Sınav süresi pek yeterli değildi. Sınavdan önce alıştıma yapılmalıydı. (28)

Yüzyüze Konuşma Sınavı için Öneriler (Suggestions for the FTFsa)

Bu tür sınavların tek öğretmenle yapılması beni daha rahat hissettirebilirdi (2)

Sınavdan ziyade sınıftaki katılım baz alınarak not verilse daha iyi olur bence (5)

Sınav esnasında konuşma seviyemizin bilgilerimize oranla ne kadar düşük olduğunu anladım. Daha çok egzersiz yapmalıyız, sınav değil. (6)

Bu tür aktivitelerin daha da arttırılması taraftarıyım. (9)

Bu organizasyon daha sistemli bir şekilde daha fazla ve çeşitli sorularla tekrarlanabilir.(10)

Cevaplar arasındaki duraksamanın çok puan düşürücü olmaması iyi olur. (21)

Bilgisayar Destekli Konuşma Sınavı için Öneriler (Suggestions for the CASA)

Daha fazla sorular, tabi resimler de eklenebilir. (8)

Sınavlar speaking hocalarımızla birebir yapılsa daha verimli olabiliriz. (16)

INTERMEDIATE LEVEL

Yüzyüze Konuşma Sınavına Yönelik Olumlu Tutumlar (Positive Attitudes Towards the FTFsa)

Sınav görevlisi çok güleyüzlüydü ve bu beni çok rahatlattı (30)E

Öğretmenin tutumu çok rahatlatıcıydı. Soruları yöneltirken sınavda soru sormak yerine sohbet ediyormuş havası vardı (31) B

Konuşma ve anlık cümle kurma becerimin de ölçülebilmesi hoşuma gitti. Ayrıca sınavın her öğrenci için özel olması hoşuma gitti (32)E

Soruların akışını sevdim diyebilirim. Günlük hayattan başlayıp, biraz daha özelleşip daha sonra speaking dersinde işledğimiz konulara gelmesi rahatlatıcıydı (33)B

Soruların müfredat dahilinde olmasını sevdim(34)B

Sınavda yüzyüze konuştuğum gözetmenin pozitif davranışlarını motive edici buldum böylece daha rahat konuşabildim. (35)B

Genel olan, gerçekten biriyle diyalog oluşturulabilecek konuların olmasını sevdim (36)B

Sınavı yapan okutman çok sevimli bir bayandı. Konuşamasam bile kendimi rahat hissetmemi sağladı .(37)B

Sınavıma giren hocalar çok sevecen beni rahatlatıcı davrandılar (38)B

Hocanın güler yüzlü olması (40) E

Sınav görevlilerinin rahatlatıcı tavırları ve güler yüzlü olmaları sınavın en çok hoşuma giden yanıydı. Bilgisayar bazlı bir sınavdansa yüzyüze yapılan bir konuşma sınavının kesinlikle daha etkili olduğunu düşünüyorum. Sınav görevlileri katılımcılara bir takım kullanımları hatırlatmada ve onlara yardımcı olmaktadırlar. Ayrıca daha rahat bir ortam oluşmaktadır. (41)B

Hocalarımız güler yüzlü ve anlayışlıydı. Sorular yeterince açıktı ve gayet rahat bir sınav ortamı vardı (42)B

Soruları yönelten öğretmen oldukça ilgili ve candandı. Bu tavrı rahatlatıucıydı. .(43)B

Öğretmenlerin benle olan iletişimi ve güler yüzlü olmaları (44)E

Öğretmenin güleryüzlü karşılaması hoşuma gitti ve beni biraz rahatlattı (45)E

Az da olsa konuşma imkanı .(50)E

Hoc alar çok üstümüze gelmedi (51)E

Karşıdakinin beni dinlemesi bir de ek soruları sevdim (53)E

Öğretmenlerin çok hoşgörülü olmasını ve rahtlatmasını (54) B

Sınavı oturarak yapmak güzeldi (58)B

Sorualrın kitapla paralel olması güzeldi (61) B

Iyi bir sınav tarzıydı, tamamen olmasa da bildiklerimizi aktarabildik (62)E

Sınav gayet iyiydi. Soru sayısı ve çeşitliliği arttırılabilir. Ciddiyet hosuma gitti. (63) E

Yüzyüze Konuşma Sınavına Yönelik Olumsuz Tutumlar (Negative Attitudes Towards the FTFsa)

Bunun sınav adı altında olması beni çok gerdi. Cümleler aklımda olmasına rağmen söyleyemedim. (29) B

Sınav hakkında daha önceden bilgim olmaması ve ilk defa bu sen yapılması (30)E

Sınavda yapacaklarım karşıma daha sonra not şeklinde döneceği için gergindim ve kendimi tam anlamıyla göstermek istedim, bu gerginlik beni rahatsız etti (32)E

Sınav iki gözetmen eşliğinde oldu. Biri sürekli soru sorarken diğerinin not almasından rahatsız oldum (33)B

Heyecanımı kontrol altına alamadığım için rahatsızlık duydum (34)B

İki gözetmenle yapılan konuşmalarda biriyle sohbet şeklinde akıcı konuşurken diğer gözetmenin izlemesi, not alması rahatsız ediciydi. (36)B

Belli bir süren var ve o sürede konuşmak zorundasın. Bu bende gerginliğe sebep olduğu için söyleyebileceklerimi bile unutabiliyorum.(37)B

Kendimi çok gergin hissettim. Söyleyebileceğim şeyleri o anki gerginliğimden söyleyemedim (38)B

Birebir ders gibi yapılan sınav beni en çok rahatsız eden şeydi. Konuşma sınavları daha günlük konuşma tarzında belki daha rahat bir ortamda olmalıdır. (41)B

Söylemek istediklerimi kolayca ifade edemedim. Daha düzgün ve doğru cümleler kurabilrdim. Kolayca kelimeleri hatırlayamadım. Rahatsız ediciydi.(43)B

Sıranın gelmesini beklemekten rahatsız oldum, sınıfın yarısının başka bir gün olmasını isterdim.sınav anında rahatsız edici bir durum olmadı (44)E

Sıranın gelmesini beklemek beni gerçekten rahatsız etti ve yordu (45)E

Saatlerce sıranın gelmesini beklemekten sınavdan önce öğretmenimin gerginliği arttırıcı sözlerinden rahatsız oldum. Sınavda hiçbirseyi sevmedim. İki öğretmen ayrı ayrı sınıflarda sınav yapsaydı o kadar beklemezdik (47)E

Sınav öncesinde beklemek (48)B

Çok kötü olan sırayla alma sistemi. 5 saat bekletme kabul edilemez.(50)E

Başarılı olamayacağım endişesini duydum. (52)B

Sınav güzeldi fakat süresi daha iyi ayarlanabilirdi.Okulda 2 saat fazladan beklemek gerekti bana sıra gelmesi için (53)E

Beklemekten. Sıranın gelmesini beklemek çok gerdi beni. (54) B

Aşırı stres yapmamdan (55)B

Türkçe düşünüp İngilizceye aktaramadığım için (58)B

Aklıma kelimelerin kolayca gelmeyişi (59)E

Aklıma bir türlü doğru sözcükler gelmedi bundan rahatsız oldum (61) B

Anide n meslek uydurmam gerekti bu da beni duraksattı. Duraksama ingilizcden degildi (65)E

Bilgisayar Destekli Konuşma Sınavına Yönelik Olumlu Tutumlar (Positive Attitudes Towards the CASA)

Görsel destek ve ipuçları(32)E

Yeterince zamanımızın olmasını (44)E

Sınavın bilgisayar ortamında olması beni rahatlattı (47)E

Konuşan kişinin anlaşılır konuşması güzeldi (49)B

Kısmen de olsa İngilizce konuşabilmemiz .(50)E

Sınavda aktif konuşma olması hoşuma gitti. Konuşma becerimizi bu yolla arttırabiliriz. Bu tür bir sınavın daha sık olması lazım, 2 haftada bir gibi .(51)E

Öğretmenle değil bilgisayarla konuşmak daha rahattı. (52)B

Sorular idare ederdi. Teknik problem olmasa iyiydi (53)E

Hocaların yardımcı olmasını üstün bir sabır göstererek (54) B

Görsellik güzeldi. Sınav sıkıcılıkran uzaktı (55)B

Görsel destekler olması (57)B

Görsel destekler olmasını (61) B

Soruların tarzları iyiydi (62)E

Yeni biseylerin denenmeye calısılması (64)E

Görsel öğeler daha rahat düşünmemi sagladı (66)B

Bilgisayar Destekli Konuşma Sınavına Yönelik Olumsuz Tutumlar (Negative Attitudes Towards the CASA)

Kulaklıktan(32)E

Daha önce hiç pratik yapılamasından dolayı rahatsız oldum (33)B

Bilgisayar sisteminin çok yavaş olmasından (34)B

Facetoface sınavdaki rahatlık bu sınavda yoktu (35)B

Düşüdükleirmi bir yere kısaca not alamamak kötüydü (36)B

Aklıma fazla bir şey gelmemesinden (38)B

Düşüncelerimi organize edemedim (40) E

Sorular tamamen dersteki kalıpları kullanmaya yönelikti ve İngilizce konuşma yeteneğimizi genel anlamda ölçen bir sınav değildi Yüzyüze bir sınavı tercih ederim(41)B

Cevapları tam hatırlayıp tam ve net cevap veremedim. Rahatsız ediciydi. .(43)B

Kesik kesik konuşmamdan rahatsız oldum (44)E

Yüzyüze konuşma sınavı daha iyiydi bence. Birinin benim konuşmama cevap vermemesi sinir bozucuydu ….(46) E

Stresten dolayı söyleyeceklerimi toparlayamadım. Çevremdekilerin konuşmaları beni rahatsız etti ve tam odaklanamadım (48)B

Internetin yavaş olması tıkladığımız sayfaların geç gelmesi, başkalarının sesini duymamız. .(50)E

Sınavda cevaplarımı kaydettikten sonra diğer soruya geçmeden diğer arkadaşlarımın cevaplarını duymak rqahatsız etti . sınav soruları ölçücü ve kapsamlı değildi.(51)E

Bağlantıdaki sorunlardan (52)B

Süre kısıtlıydı. Çok kısa sürede sınavın sistemine alışmak zordu. Yanlış tıklamayla sınav kapanıypor ya da her halukarda kapanıyordu (?) bilgisayarların yenilenmesi , sınav süresinin uzatılması ve sistemin daha kolay olması lazım. (53)E

İlk kez böyle bir sınava gridiğim için gergindim. Önceki günlerde alıştırmalar falan olsa daha iyi olurdu. (54) B

Bağlantıdaki sorunlardan rahatsız oldum ve asırı heyecan yapmıstım (55)B

Zaman azdı (57)B

İlk defa böyle bir sınava girmenin heyecanı vardı (58)B

Bazı seyleri anlamadım (61) B

Bilgisayarlar yetersizdi. Bu yüzden tam performans gösteremedim . bilgisayarlar yenilenmeli. Daha organize bir sistem kurulmalı. (62)E

Bilgisayara cevap vermek gercekten rahatsz ediciydi . cevapların kaydedilip edilmediginin farkında degildim. (63) E

Internetin yavaslıgı ve programın karısıklıgı benden istenen soruları yapamamama neden oldu (65) E

Kendi kendime konumsak ve ses kaydı rahatsz etti (66)B

Yüzyüze Konuşma Sınavı için Öneriler (Suggestions for the FTFsa)

Değerlendirme işlemi sınavdan sonra yapılabilirdi. (33)B

Konuşma sınavı yapmak yerine konuşmayla ilgili pratikler yapılsa bence daha yararlı olabilir. Kendimizi gergin hissetmeyiz en azından. (38)B

Bu etkinliklerin sınav bazından çıkıp derslerde yaptığımız aktivitelere dönüşmesi daha faydalı olacaktır. Ders kitapları ile çok da yararlı bir speaking dersi işleyemiyoruz. (39)B

APPENDIX K: Foreign Language Speaking Test Anxiety and Speaking Anxiety Questionnaire

To the attention of the participants,

Composed of two sections, this questionnaire has been prepared to gather information about your foreign language speaking test anxiety and speaking anxiety. The information obtained from the questionnaire will be used to support a research study conducted at MA TEFL program at Bilkent University. There are no correct or wrong answers in this questionnaire. Please mark the option that best reflects you and respond to all questions, this is rather important for the validity and reliability of the questionnaire. The responses you give to the questions will be kept strictly confidential.

I would like to remind that your responses are quite valuable to the study and I thank you in advance for your time and effort.

Ebru Öztekin
Graduate student
MA TEFL, Bilkent University

Informed consent form:

I understood the content and purpose of the questionnaire. I agree to complete the questionnaire and to let the researcher use my responses in the scientific study on condition that my information is kept confidential.

Participant name/ surname: _____ Signature:

Class: : _____

Gender/ Age : _____

Date :_____

This questionnaire is composed of three pages and two sections.
The options in the questionnaire are as below:

Strongly agree
Agree
Undecided
Disagree
Strongly disagree

Please go to the next page to start answering the questions.

FIRST SECTION: **Please mark the option that best describes you for each statement.**	**Strongly agree**	**Agree**	**Undecided**	**Disagree**	**Strongly disagree**
1. I feel tense in the speaking test.					
2. I would like to go to the rest room during the speaking test.					
3. My heart starts pounding in the speaking test.					
4. I feel irritated in the speaking test.					
5. I respond to all questions in the speaking tests consciously.					
6. I feel thirsty in the speaking test.					
7. I sleep comfortably the day before the speaking test.					
8. I yawn in the speaking test.					
9. My hands sweat in the speaking test.					
10. I have difficulty in organizing my thoughts in the speaking test.					
11. I feel sleepy during the speaking test.					
12. I feel free to talk in the speaking test.					
13. I feel confident in the speaking test.					
14. I am unaware of what I do in the speaking test.					
15. I sweat in the speaking test.					

16. My mind goes blank in the speaking test.					
17. My voice tremble in the speaking test.					
18. I feel relaxed in the speaking test.					
19. I have a sore shoulder in the speaking test.					
20. I cannot concentrate in the speaking test.					
21. I wonder how other students performed in the test when I am in the speaking test.					
22. During the speaking test, I think that I would be successful.					
23. I feel very comfortable if I get prepared for the speaking test beforehand.					

SECOND SECTION: **Please mark the option that best describes you for each statement.**	**Strongly agree**	**Agree**	**Undecided**	**Disagree**	**Strongly disagree**
1. I never feel confident when I speak in English classes.					
2. I am not scared of making mistakes while speaking in English classes.					
3. I tremble when I understand that my name would be called in English classes.					
4. I still think that other students are better than me in speaking English.					
5. I start to panic when I am asked to speak in English classes without preparation.					
6. I don't understand why some people are so uncomfortable with speaking in English classes.					

7. I feel so anxious when I am asked to speak in English classes that I even forget what I know.					
8. I feel self conscious when it comes to volunteering to answer a question in English classes.					
9. I feel anxious about speaking in class even if I get prepared for the English class at home.					
10. I feel confident when I speak in English classes.					
11. When I am called on in English classes, I feel that my heart stands still.					
12. I don't feel the necessity to get prepared very well to speak in the English classes.					
13. I feel ashamed of speaking English in front of other students.					
14. English classes move so quickly that I worry about getting left behind.					
15. I feel anxious and confused as I speak in English classes.					
16. I feel overwhelmed by the number of rules you have to learn to speak English.					
17. I am afraid other students will laugh at me when I speak English.					
18. I get nervous when my English teacher asks questions which I haven't prepared in advance.					
19. My mind would go blank when I am asked to speak English anywhere.					
20. My speech will be incoherent if I speak in English in anywhere.					
21. I can impress the audience with my ability to speak English in anywhere.					
22. I will get tongue-tied if someone asks me to speak English in anywhere.					
23. I feel very comfortable if I practice speaking English when I am alone.					

24. I can easily organize my thoughts in English if I practice speaking English when I am alone.	☐	☐	☐	☐	☐
25. I can easily organize my thoughts in English if I practice speaking English when I am with my close friends.	☐	☐	☐	☐	☐

Is there anything you want to add?

_____

_____

_____

This is the end of the questionnaire. Thanks again for your participation ☺

Contact information for your questions and suggestions:

E-mail: educationline.tr@gmail.com

APPENDIX L: Yabancı Dil Konuşma Sınavı Kaygısı Ve Konuşma Kaygısı Ölçeği

Katılımcıların dikkatine,

İki bölümden oluşan bu anket, yabancı dil konuşma sınavına ve yabancı dilde konuşmaya yönelik kaygı düzeyiniz hakkında bilgi toplamak için hazırlanmıştır. Anketten elde edilen bilgiler Bilkent Üniversitesi Yabancı Dil Olarak İngilizce Öğretimi Yüksek Lisans (MA TEFL) Bölümünde yapılmakta olan bir araştırmanın içeriğine destek olması amacıyla kullanılacaktır. Bu ankette doğru ya da yanlış cevaplar yoktur. Lütfen soruları sizi en iyi yansıtan şıkkı seçerek işaretleyiniz ve tüm soruları cevaplayınız; bunu yapmanız anketin geçerliliği ve güvenilirliği açısından oldukça önemlidir. ***Anket*** sorularına verdiğiniz cevaplar kesinlikle ***gizli tutulacaktır*.**

Vereceğiniz cevapların bu çalışma için çok değerli olduğunu hatırlatarak ayırdığınız zaman ve emeğiniz için şimdiden çok teşekkür ederim.

Ebru Öztekin
Yüksek lisans öğrencisi
MA TEFL, Bilkent Üniversitesi

Aydınlatılmış onam formu:

Anketin içeriğini ve amacını anladım. Anketi cevaplamayı ve bilgilerimin gizli tutulması şartıyla cevaplarımın ilgili bilimsel çalışmada kullanılmasını kabul ediyorum.

Katılımcının adı/ soyadı: _____ İmza:

Sınıfı/Şubesi: **:** _____

Cinsiyet/ Yaş **:** _____

Tarih **:**_____

Bu anket üç sayfadan ve iki bölümden oluşmaktadır.
Anketteki seçenekler aşağıdaki gibidir:

Kesinlikle katılıyorum
Katılıyorum
Kararsızım
Katılmıyorum
Kesinlikle katılmıyorum

Ankete başlamak için lütfen bir sonraki sayfaya geçiniz.

BİRİNCİ BÖLÜM: **Lütfen her bir soru için sizi en iyi yansıtan seçeneği işaretleyin.**	**Kesinlikle katılıyorum**	**Katılıyorum**	**Kararsızım**	**Katılmıyorum**	**Kesinlikle katılmıyorum**
24. Konuşma sınavında gerilirim.	☐	☐	☐	☐	☐
25. Konuşma sınavında lavaboya veya tuvalete gidesim gelir.	☐	☐	☐	☐	☐
26. Konuşma sınavında kalbim çok hızlı atmaya başlar.	☐	☐	☐	☐	☐
27. Konuşma sınavında tedirgin hissederim.	☐	☐	☐	☐	☐
28. Konuşma sınavında tüm soruları bilinçli bir şekilde cevaplarım.	☐	☐	☐	☐	☐
29. Konuşma sınavında su içme ihtiyacı duyarım.	☐	☐	☐	☐	☐
30. Konuşma sınavından önceki gece çok rahat uyurum.	☐	☐	☐	☐	☐
31. Konuşma sınavında esnerim.	☐	☐	☐	☐	☐
32. Konuşma sınavında ellerim terler.	☐	☐	☐	☐	☐
33. Konuşma sınavında düşüncelerimi organize etmekte güçlük çekerim.	☐	☐	☐	☐	☐
34. Konuşma sınavında uykulu hissederim.	☐	☐	☐	☐	☐
35. Konuşma sınavında kendimi konuşma konusunda özgür hissederim.	☐	☐	☐	☐	☐
36. Konuşma sınavında özgüvenli hissederim.	☐	☐	☐	☐	☐
37. Konuşma sınavında ne yaptığımın farkında olmam.	☐	☐	☐	☐	☐
38. Konuşma sınavında terlerim.	☐	☐	☐	☐	☐

39. Konuşma sınavında hafıza kayıpları yaşarım.					
40. Konuşma sınavında sesim titrer.					
41. Konuşma sınavında rahat hissederim.					
42. Konuşma sınavında omuzlarım tutulur.					
43. Konuşma sınavında konsantre olamam.					
44. Konuşma sınavında diğer öğrencilerin nasıl bir performans gösterdiklerini düşünürüm.					
45. Konuşma sınavı sırasında sınavda başarılı olacağımı düşünürüm.					
46. Konuşma sınavına önceden iyi hazırlandıysam kendimi çok rahat hissederim.					

İKİNCİ BÖLÜM: **Lütfen her bir soru için sizi en iyi yansıtan seçeneği işaretleyin.**	**Kesinlikle katılıyorum**	**Katılıyorum**	**Kararsızım**	**Katılmıyorum**	**Kesinlikle Katılmıyorum**
26. İngilizce derslerinde konuşurken asla kendimden emin olamam.					
27. İngilizce derslerinde konuşurken hata yapmaktan korkmam.					
28. İngilizce derslerinde konuşmam için bana seslenileceğini anladığımda titrerim.					
29. Hala diğer öğrencilerin İngilizce konuşmada benden daha iyi olduklarını düşünüyorum.					
30. İngilizce derslerinde ön hazırlıksız olarak konuşmak gerektiğinde paniklemeye başlıyorum.					
31. Bazı insanların İngilizce derslerinde konuşmaktan neden bu kadar rahatsız olduklarını anlamıyorum.					

32. İngilizce derslerinde benden konuşmam istendiğinde o kadar geriliyorum ki, bildiğim şeyleri unutuyorum.					
33. İngilizce derslerinde bir şeye cevap vermek için gönüllü olmaya utanıyorum.					
34. İngilizce dersine iyi hazırlanmış olsam bile sınıfta konuşma konusunda kaygı duyuyorum.					
35. İngilizce dersinde konuşurken kendimden emin hissediyorum.					
36. İngilizce dersinde bana seslenildiğinde kalbim duracak gibi oluyor.					
37. İngilizce dersinde konuşmak için çok iyi hazırlanma zorunluluğu hissetmiyorum.					
38. Diğer öğrencilerin önünde İngilizce konuşmaya utanırım.					
39. İngilizce dersleri o kadar hızlı ilerliyor ki, geride kalmaktan korkuyorum.					
40. İngilizce derslerinde konuşurken geriliyorum ve kafam karışıyor.					
41. İngilizce konuşmak için öğrenmem gereken kuralların sayısı altında eziliyorum.					
42. Korkarım ki İngilizce konuştuğumda diğer öğrenciler bana güler.					
43. İngilizce öğretmeni cevaplamaya önceden hazırlanmadığım sorular sorduğunda gerilirim.					
44. Herhangi bir yerde benden İngilizce konuşmam istendiğinde beynim durur.					
45. Herhangi bir yerde İngilizce konuşursam, konuşmam tutarsız olur.					
46. Herhangi bir yerde, beni dinleyicileri İngilizce konuşma yeteneğimle etkileyebilirim.					
47. Dışarıda birisi benden İngilizce konuşmamı isterse dilim dolanır.					
48. Yalnız başıma İngilizce konuşma pratiği yaparken kendimi çok rahat hissederim.					

49. Yalnız başıma İngilizce konuşma pratiği yaparken düşüncelerimi kolaylıkla organize edebilirim.	☐	☐	☐	☐	☐
50. Yakın arkadaşlarımın yanında İngilizce konuşma pratiği yaparken düşüncelerimi kolaylıkla organize edebilirim.	☐	☐	☐	☐	☐

Eklemek istediğiniz başka bir şey var mı?

_____

_____

_____

Anketiniz bitmiştir. Katılımınız ve katkınız için tekrar teşekkürler ☺

Soru ve önerileriniz için iletişim adresi: E-posta: educationline.tr@gmail.com

APPENDIX M: Computer Attitudes Questionnaire

To the attention of the participants,

This questionnaire has been prepared to shed light into your attitudes towards computers and using computers. The information obtained from the questionnaire will be used to support a research study conducted at MA TEFL program at Bilkent University. There are no correct or wrong answers in this questionnaire. Please mark the option that best reflects you and respond to all questions, this is rather important for the validity and reliability of the questionnaire. The responses you give to the questions will be kept strictly confidential.

I would like to remind that your responses are quite valuable to the study and I thank you in advance for your time and effort.

Ebru Öztekin
Graduate student
MA TEFL, Bilkent University

Informed consent form:

I understood the content and purpose of the questionnaire. I agree to complete the questionnaire and to let the researcher use my responses in the scientific study on condition that my information is kept confidential.

Participant name/ surname: _____ Signature:

Class: : _____

Gender/ Age : _____

Date :_____

This questionnaire is composed of two pages and 30 questions.
The options in the questionnaire are as below:

Strongly agree
Agree
Undecided
Disagree
Strongly disagree

Please go to the next page to start answering the questions.

Please choose the option that best describes you for each statement.

	Strongly agree	Agree	Undecided	Disagree	Strongly disagree
1. I do ANYoperation on computers with ease.					
2. Anyone can learn to use computers very effectively if they want it.					
3. Except for the internet, I can use medium-level programs with all their properties, too. (i.e.: Excel, Word, virüs programs, compression utilities)					
4. It is easy for me to develop computer software.					
5. I can resolve the problem by myself if my computer breaks down.					
6. Computers are my best friends.					
7. People are gradually becoming slaves of the Internet					
8. I can get good grades even in a difficult computer course.					
9. I don't work on the computer if it is possible to complete a task in some other way.					
10. The Internet is dehumanizing people.					
11. I don't think that I can learn a computer programming langauge.					
12. The internet saves us from a lot of tiring stuff.					
13. I get nervous when I need to do something on the computer.					
14. Working on computers is easier than doing things without them.					
15. I can easily follow the advancements in the world of computers.					
16. I have difficulty in understanding the technical details about computers.					
17. You have to be a genius even to find the links to click on some websites.					
18. I am scared of losing a lot of data by pushing the wrong button.					

19. Computers are an indispensible part of education and business life.					
20. I think working on the computer is rather enjoyable.					
21. I wish computers were not this involved in our lives.					
22. I find working on the computer very boring.					
23. Our life speeds up and gets easier thanks to the Internet.					
24. I hate computers.					
25. It is difficult to use the Internet effectively.					
26. I can easily create and organize text and presentations on the computer.					
27. It is not easy for me to record my voice on the computer.					
28. I panic when I am to try something new on the computer.					
29. It is possible to have the same effectiveness in education and work environments without computers.					
30. I look forward to using computers for any work of mine.					

Is there anything you want to add?

_____

_____

_____

This is the end of the questionnaire. Thanks again for your participation ☺

Contact information for your questions and suggestions:

E-mail: educationline.tr@gmail.com

APPENDIX N: Bilgisayar Tutum Ölçeği

Katılımcıların dikkatine,

Bu anket bilgisayarlara ve bilgisayar kullanımına bakış açınıza ışık tutmak için hazırlanmıştır. Anketten elde edilen bilgiler Bilkent Üniversitesi İngilizce Öğretimi Yüksek Lisans (MA TEFL) Bölümünde yapılmakta olan bir araştırmanın içeriğine destek olması amacıyla kullanılacaktır. Bu ankette doğru ya da yanlış cevaplar yoktur. Lütfen soruları size en yakın gelen şıkkı seçerek işaretleyiniz ve tüm soruları cevaplayınız; bunu yapmanız anketin geçerliliği ve güvenilirliği açısından oldukça önemlidir. Anket sorularına verdiğiniz cevaplar kesinlikle gizli tutulacaktır.

Vereceğiniz cevapların bu çalışma için çok değerli olduğunu hatırlatarak ayırdığınız zaman ve emeğiniz için şimdiden çok teşekkür ederim.

Ebru Öztekin
Yüksek lisans öğrencisi
MA TEFL, Bilkent Üniversitesi

Aydınlatılmış onam formu:

Anketin içeriğini ve amacını anladım. Anketi cevaplamayı ve bilgilerimin gizli tutulması şartıyla cevaplarımın ilgili bilimsel çalışmada kullanılmasını kabul ediyorum.

Katılımcının adı/ soyadı : _____

Sınıfı/Şubesi: : _____

Cinsiyet/ Yaş : _____

Tarih :_____

İmza:

Bu anket iki sayfadan ve 30 sorudan oluşmaktadır.
Anketteki seçenekler aşağıdaki gibidir:

Kesinlikle katılıyorum
Katılıyorum
Kararsızım
Katılmıyorum
Kesinlikle katılmıyorum

Ankete başlamak için lütfen bir sonraki sayfaya geçiniz.

Lütfen her bir soru için sizi en iyi yansıtan seçeneği işaretleyin.

	Kesinlikle katılıyorum	Katılıyorum	Kararsızım	Katılmıyorum	Kesinlikle katılmıyorum
1. Bilgisayarda istediğim HER işlemi rahatlıkla yaparım.					
2. İsteyen herkes çok iyi bilgisayar kullanmayı öğrenebilir.					
3. İnternet dışında, orta düzey programları da tüm özellikleriyle kullanabilirim. (ör: Excel, Word, virüs tarama, dosya sıkıştırma)					
4. Bilgisayar programlarının yazılımlarını geliştirmek benim için kolaydır.					
5. Bilgisayarım arızalansa sorunu tek başıma halledebilirim.					
6. Bilgisayar benim en iyi arkadaşımdır.					
7. İnsanlar gittikçe internetin kölesi haline geliyor.					
8. Zor bir bilgisayar dersinde bile iyi notlar alabilirim.					
9. Eğer başka bir şekilde halletmem mümkünse, işimi bilgisayarda yapmam.					
10. İnternet insanları makineleştiriyor.					
11. Herhangi bir bilgisayar programlama dili öğrenebileceğimi düşünmüyorum.					
12. İnternet bizi yorucu birçok işten kurtarır.					
13. Bilgisayarda yapmam gereken bir iş olduğunda gerilirim.					
14. Bilgisayarda yaptığım çalışmalar bana daha kolay gelir.					
15. Bilgisayar dünyasında gerçekleşen yenilikleri rahatlıkla takip edebilirim.					
16. Bilgisayarlarla ilgili teknik detayları anlamakta zorlanıyorum.					
17. Bazı web sitelerinde tıklanacak yerleri bulmak için bile dahi olmak gerekiyor.					
18. Bilgisayarda yanlış bir tuşa basarak bir sürü bilgi kaybetmek beni korkutuyor.					

19. Bilgisayarlar eğitim ve iş hayatının vazgeçilmez birer parçasıdır.					
20. Bilgisayarda iş yapmanın çok eğlenceli olduğunu düşünüyorum.					
21. Keşke bilgisayarlar hayatımıza bu kadar dahil olmasaydı.					
22. Bilgisayarda çalışmak bana çok sıkıcı gelir.					
23. İnternet sayesinde hayatımız hızlanıyor ve kolaylaşıyor.					
24. Bilgisayarlardan nefret ederim.					
25. İnterneti etkin bir biçimde kullanmak zordur.					
26. Bilgisayarda kolaylıkla metin ve sunum hazırlayabilir ve düzenleyebilirim.					
27. Sesimi bilgisayara kaydetmek benim için kolay bir şey değildir.					
28. Bilgisayarda yeni bir şey denemem gerektiğinde paniklerim.					
29. Eğitim ve çalışma ortamlarında bilgisayarlar olmasa da aynı verim alınabilir.					
30. Herhangi bir çalışmamda bilgisayar kullanmayı dört gözle beklerim.					

Eklemek istediğiniz başka bir şey var mı?

_____

_____

_____

Anketiniz bitmiştir. Katılımınız ve katkınız için tekrar teşekkürler ☺

Soru ve önerileriniz için iletişim adresi: E-posta: educationline.tr@gmail.com