

THE EFFECTIVENESS OF CLOZE TESTS IN ASSESSING THE
SPEAKING/WRITING SKILLS OF UNIVERSITY EFL LEARNERS

A Master's Thesis

by

UFUK BİLKİ

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

July 2011

To My Beloved Family

THE EFFECTIVENESS OF CLOZE TESTS IN ASSESSING THE
SPEAKING/WRITING SKILLS OF UNIVERSITY EFL LEARNERS

The Graduate School of Education

of

Bilkent University

by

UFUK BİLKİ

In Partial Fulfillment 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

BILKENT UNIVERSITY
THE GRADUATE SCHOOL OF EDUCATION
MA THESIS EXAMINATION RESULT FORM

July 7 2011

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

Ufuk Bilki

has read the thesis of the student.

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

Thesis Title: The Effectiveness of Cloze Tests in Assessing the
Speaking/Writing Skills of University EFL Learners.

Thesis Advisor: Asst. Prof. Dr. Philip Durrant
Bilkent University, MA TEFL Program

Committee Members: Asst. Prof. Dr. JoDee Walters
Bilkent University, MA TEFL Program

Dr. William Snyder
Columbia University, Teachers College

ABSTRACT

THE EFFECTIVENESS OF CLOZE TESTS IN ASSESSING THE
SPEAKING/WRITING SKILLS OF UNIVERSITY EFL LEARNERS.

Ufuk Bilki

M.A. The Program of Teaching English as a Foreign Language

Supervisor: Asst. Prof. Dr. Philip Durrant

July 2011

This study investigated the effectiveness of cloze tests in assessing the speaking and writing skill levels of Manisa Celal Bayar University Preparatory School students. The study also tried to find out whether there are differences in the success levels of such tests in assessing speaking and writing skill levels when the conditions are changed. It examined different text selection, deletion methods and scoring methods to determine whether there are any differences in the success levels.

The study was conducted in Celal Bayar University Preparatory School with 60 intermediate (+) students from three different classes. The students were tested using six different cloze tests. The cloze tests differed in text selection and deletion methods. These cloze tests were then scored using two different scoring methods. The scores were

correlated with the speaking and writing scores of the students to check whether there is a correlation between the cloze test results and the students' writing/speaking scores. The correlations were also compared to each other to find if there are differences in the success levels of the tests' ability to assess the different skills.

The results pointed out a possible use of cloze tests in testing speaking/writing. Although strong correlation was found between writing/speaking tests and cloze tests, the numbers indicate that the cloze tests might not be adequate measuring tools for the productive skills on their own. This suggests a supportive use of cloze tests. The differing correlations also suggested a careful use of such tests. In conclusion, cloze tests can be valuable tools, if used correctly, to help assess speaking and writing skills.

These results mostly mean that more studies should be conducted on the subject. The differences between the correlations of different preparation and scoring methods suggest that there might be some chance to further refine the cloze tests to make them more appropriate as alternative testing methods.

Key words: cloze procedure, EFL, testing, speaking skills, writing skills, indirect testing, text selection, deletion methods, scoring methods

ÖZET

ÇIKARTMALI SINAVLARIN ÜNİVERSİTE DÜZEYİNDEKİ YABANCI DİL OLARAK İNGİLİZCE ÖĞRENEN ÖĞRENCİLERİN KONUŞMA VE YAZMA BECERİLERİNİ ÖLÇME KONUSUNDAKİ ETKİNLİĞİ

Ufuk Bilki

Yüksek Lisans, Yabancı Dil Olarak İngilizce Öğretimi Programı

Tez Yöneticisi: Yrd. Doç. Dr. Philip Durrant

Temmuz 2011

Bu çalışma çıkartmalı sınavların Manisa Celal Bayar Üniversitesi Hazırlık Okulu öğrencilerinin konuşma ve yazma becerilerini ölçme konusundaki etkinliğini konu almıştır. Çalışma aynı zamanda, bu tarz sınavların nitelikleri değiştirildiğinde konuşma ve yazma becerilerini ölçme etkinliklerinde değişiklik olup olmadığını da incelemiştir. Bu amaçla farklı metin seçimleri, boşluklama yöntemleri, ve puanlandırma sistemleri incelenmiştir.

Çalışmaya Celal Bayar Üniversitesi Hazırlık Okulu'nda okuyan, üç ayrı sınıftan 60 orta (ve üstü) seviye öğrenci katılmıştır. Katılımcılar altı ayrı çıkartmalı sınav kullanılarak test edilmiştir. Çıkartmalı sınavlar metin seçimleri ve boşluklama

yöntemleri açısından birbirinden farklıdır. Bu sınavlar daha sonra iki farklı puanlandırma sistemi kullanılarak değerlendirilmiştir. Sonuçlar öğrencilerin konuşma ve yazma sınavlarından aldıkları puanlarla ilintilendirilmiş ve öğrencilerin çıkartmalı sınav sonuçlarıyla yazma/konuşma sınavı sonuçları arasında bir korelasyon olup olmadığı gözlenmiştir. Korelasyon sonuçları aynı zamanda birbirleri ile kıyaslanarak testlerin becerileri ölçmedeki başarıları arasındaki farklar da incelenmiştir.

Sonuçlar çıkartmalı sınavların, öğrencilerin konuşma ve yazma becerilerinin ölçülmesinde kullanılabileceğini göstermiştir. Her ne kadar konuşma/yazma sınav sonuçları ile çıkartmalı test sonuçları arasında güçlü bir korelasyon görülmüş olsa da, sonuçlar çıkartmalı sınavların tek başlarına konuşma ve yazma becerilerini ölçme konusunda yeterli olmayacağını göstermiştir. Bu durum bizlere çıkartmalı sınavların daha çok destek amaçlı kullanılabilir olduğunu göstermiştir. Üstelik, çeşitli nitelik değişiklikleri sonucunda farklılık gösteren korelasyonlar da bizlere bu tür testlerin kullanımında dikkatli olmamız gerektiğini söylüyor. Sonuç olarak, çıkartmalı testlerin doğru kullanılırlarsa öğrencilerin konuşma ve yazma becerilerini ölçmede değerli araçlar olabileceğini görüyoruz.

Bu sonuçlar genel olarak konu üzerinde daha fazla çalışma yapılması gerektiğini gösteriyor. Korelasyonlar arasındaki farklı hazırlama ve puanlama teknikleri dolayısı ile oluşan farklar, çıkartmalı sınavların daha da rafine edilerek alternatif değerlendirme tekniklerine dönüştürülebilecek olduklarını gösteriyorlar.

Anahtar Kelimeler: çıkartmalı sınavlar, yabancı dil olarak İngilizce eğitimi ölçme ve değerlendirme, konuşma becerileri, yazma becerileri, endirekt testler, boşluklama yöntemleri, puanlandırma sistemleri

ACKNOWLEDGMENTS

There are a few people I would like to thank for their help and support throughout my MATEFL experience of 2010-2011. To begin with, I would like to thank to the MATEFL staff that had been trying to help me achieve. I would like to thank my thesis advisor Dr. Philip Durrant for his support, guidance, and precious feedback. I also would like to thank to my other teachers in the program, Dr. Maria Angelova, Dr. Julie Mathews-Aydınlı, and Dr. JoDee Walters for their efforts during the year.

I would also like to express my profound love and gratitude to my family. I feel indebted to my best friend and spouse Emel, my parents Meral and Mustafa and my sister Fulya who had given their continuous love, patience, and support during the semester and has always been there for me.

Moreover, I would like to thank to Dr. Ayça Ülker Erkan and Dr. Münevver Büyükyazı, and the rest of my colleagues in Celal Bayar University who had made it possible for me to attend to this program and who had supported me throughout the year.

Finally, I would like to express my gratitude to my classmate Zeynep Erşin who had always been there for me throughout the year.

TABLE OF CONTENTS

ABSTRACT	iv
ÖZET.....	vi
ACKNOWLEDGMENTS.....	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xiv
CHAPTER 1: INTRODUCTION	1
Introduction.....	1
Background of the Study.....	2
Statement of the Problem.....	6
Research Questions	7
Significance of the Study	8
Conclusion.....	9
CHAPTER II: LITERATURE REVIEW	10
Introduction.....	10
Testing Writing	10
Testing Speaking.....	12
Indirect Testing	14
Studies on Cloze Tests	16
What are cloze tests.....	16
Scoring methods.....	17
What do cloze tests measure	23

Conclusion.....	30
CHAPTER III: METHODOLOGY	32
Introduction.....	32
Setting.....	33
How is Writing Taught and Assessed in CBU?	34
How is Speaking Taught and Assessed in CBU?.....	35
Participants	36
Instruments	37
The Cloze tests	37
The Article Type of Cloze Tests	38
The Dialogue Type of Cloze Tests.....	39
Scoring The Cloze Tests	40
The Speaking Test.....	41
The Writing Test	42
Procedure.....	43
Data Analysis	45
Conclusion.....	45
CHAPTER IV: DATA ANALYSIS	46
Overview	46
Data Analysis Procedure	47
The Results of the Six Cloze Tests.....	48
Decision on Normality	54

Correlations of Cloze Tests with Writing and Speaking Tests	54
General Correlation of the Writing/Speaking Tests with the Whole Group's Cloze Test Scores	55
Correlation of the Writing/Speaking Tests with the Different Groups of Cloze Tests	56
Correlations within Groups – Article Text Type	56
Correlations within Groups – Dialogue Text Type	58
Standalone Parametric Correlations Where Possible	60
Conclusion.....	62
CHAPTER V: CONCLUSION.....	63
Introduction	63
Results and Discussion.....	64
Research Question 1: To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' speaking skills?	64
Cloze Test Results' Correlated as a Single Group	64
Cloze Test Results Correlated Contrastively between Deletion Method Groups	65
Research Question 2: To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' writing skills?	67
Cloze Test Results' Correlated as a Single Group	67
Cloze Test Results Correlated Contrastively between Deletion Method Groups	68

Research Question 3: Are there important differences between cloze test preparation and scoring methods with regard to how well they assess students' speaking and writing skills?	71
Limitations of the Study	72
Pedagogical Implications of the Study.....	73
Suggestions for Further Research	75
Conclusion.....	76
REFERENCES	77
APPENDIX A: UN-MUTILATED TEXTS	80
APPENDIX B: THE SIX CLOZE TESTS	86
APPENDIX C: ACCEPTABLE ANSWERS	104

LIST OF TABLES

Table 1: Descriptive Statistics for the Cloze Tests	49
Table 2: Normality Tests for Cloze Test Results	50
Table 3: Descriptive Statistics for the Writing and Speaking Tests.....	51
Table 4: Normality Tests for Writing and Speaking Tests.....	52
Table 5: Normality Tests for Writing and Speaking Tests (Groups)	53
Table 6: Writing and Speaking Tests Correlated with Cloze Scores (Groups Together)	55
Table 7 – Writing and Speaking Tests Correlated with Cloze Scores (For groups of Article Type Text).....	57
Table 8 – Writing and Speaking Tests Correlated with Cloze Scores (For groups of Dialogue Type Text)	59
Table 9 – Parametric Correlations within Possible Groups: Writing and Speaking Tests Correlated with Cloze Scores in groups according to deletion methods.	61

CHAPTER 1: INTRODUCTION

Introduction

Assessing the skills of EFL/ESL learners has always been an important part of language teaching. According to Oller Jr. (1974) the basic principle behind all language skills is the ability to predict the flow of the elements. The best test to assess such ability would be a contextualized gap filling test. Such a gap filling test, the cloze test, is the main tool that was used in this thesis.

The cloze procedure, which was first used in language teaching by Wilson Taylor in 1953, has proven to be a very valuable tool in many different situations. A great number of researchers have found cloze tests to be useful in testing the listening, reading, vocabulary, and grammar skills of language learners. However, the effectiveness of the cloze procedure in assessing the productive skills of speaking and writing has not been given enough attention.

Today, some institutions in Turkey assess the language levels of students by only measuring their grammatical and vocabulary knowledge in addition to their reading and listening skills. Such institutions overlook measuring speaking and writing skills most of the time as they are very time consuming to test and score. Meanwhile, some of the bigger institutions (for example universities) try to assess the speaking and writing skills of their students with arduous and impractical methods which sometimes may be inadequate to fully test the speaking and writing potentials of their students. Such methods (for example, interviews, portfolios, or essay writing) require so much time and

effort that reliable periodic testing of speaking and writing becomes virtually impossible for some institutions.

This research tries to find out whether (and how) cloze tests can contribute to the assessment of speaking and writing skills of students reliably and, most importantly, practically.

Background of the Study

The use of passage completion tasks in psychological testing appears to have originated from the German psychologist Hermann Ebbinghaus (1850-1909), who first reported on his new procedure in 1897 (as cited in Harris, 1985). More than half a century later, cloze tests were introduced to language studies by Wilson Taylor in 1953 as a procedure to assess the readability of written texts (Taylor, 1953). The word 'cloze' is derived from the gestalt concept of closure - completing a pattern (Merriam-Webster Online Dictionary (www.m-w.com); Stansfield, 1980). In the cloze procedure, every nth word of a passage is systematically deleted and the reader is asked to fill in the missing word. Later, to better shape the technique according to the needs of the testers, researchers suggested removing words selectively at reasonable intervals instead of consistently deleting every "nth" word from the text (Wipf, 1981). Steinman (2002) suggests that the procedure can be applied to any language.

Since its debut in the early 50s as a language testing device, the cloze procedure has gathered much attention and controversy. The cloze procedure has been used and found to be an adequate measurement device in a very wide area of subjects ranging from the investigation of social class differences, to the assessment of general language

proficiency. On the other hand, language testing researchers have argued not only about what cloze tests actually measure but also about many different scoring and deletion methods over the years.

One of the biggest research areas regarding cloze tests is on how to score them. Oller Jr. (1972) worked with a great number of subjects to test five different scoring methods including the exact word scoring method and acceptable word scoring method (these methods will be further discussed later in the literature review). The research was conducted on approximately 400 students from 55 different native language backgrounds and Oller stated that all five scoring techniques provided similar results. Clausing and Senko (1978) suggested the use of two different scoring methods other than the exact word method. Their 'Three-Stage Scoring Hierarchy for Partial Credit' and 'Quick Performance Measurement and Feedback Technique' showed promising results. Brown (1980) also found similar results for four different scoring methods. All these different methods will be discussed in Chapter Two.

Cloze tests have mainly been used to assess language competence and/or different skill levels of second/foreign language learners. Stubbs and Tucker (1974) used the procedure as a measure of English proficiency. They finished their enthusiastic report by stating that the cloze test is a powerful and economical tool to test the English language proficiency of non-native speakers. They also regarded cloze tests as useful diagnostic tools for teachers. Darnell (1968) conducted another study on the usage of cloze tests as general English proficiency measurement tools. In his study, he correlated the cloze test results of 48 students with their TOEFL scores. He found a very high

positive correlation between the two scores. Yamashita (2003) tried to find out whether cloze tests can test reading by having skilled and less skilled readers take the same test. The results not only showed that cloze tests are a good measure of reading skills but also revealed that cloze tests are completed by both skilled and less skilled readers using text-level information. This result shows that cloze tests are not merely fill-in-the-blank type exercises, and students need a real understanding of the text to be successful in a cloze test. Hadley and Naaykens (1997) compared the cloze test scores of their students with a more traditional grammar test. They found that the cloze procedure correlated very highly with the grammar test. Although most of the studies on cloze tests have been conducted in English there has also been some research in other languages. For example, Brière, Clausen, Senko, and Purcell, (1978) conducted an investigation on the use of cloze tests in measuring native English speakers' achievement in four different foreign languages: German, Japanese, Russian and Spanish. Their study showed that cloze tests are sensitive enough to separate the achievement scores of first, second and third semester students of those languages.

Even though there have been many studies done on using cloze tests to assess various skills of language learners, researchers have generally overlooked the topic of assessing writing and speaking skills with the use of the cloze procedure. Hanania and Shikhani (1986) conducted a large study at the American University of Beirut comparing the results of cloze tests, a proficiency exam and a 250 word writing task. Out of the 1572 participants, 337 were selected as a sample group to carry out the correlations. The researchers reported a high correlation of cloze tests with both the

proficiency exam and the writing task. While they recommended further research, their findings seemed interesting. However, the 250 word limit on the writing test does not seem to fully reflect the writing proficiency expected from a university level student. There were also some studies that attempted to measure speaking skills with cloze tests. Shohamy (1982) investigated the correlation between a Hebrew Oral Interview speaking test and the cloze procedure. The study had 100 participants who were enrolled in Hebrew classes in the University of Minnesota. Shohamy used two different cloze tests in addition to an oral interview in Hebrew for her study. The results showed that there is a very high positive correlation between cloze test scores and oral interview scores of the participants. Even though this study shows that cloze tests correlate highly with oral proficiency tests it did not include any specific preparation techniques for cloze tests to test speaking skills. The two cloze tests only differed in their difficulty without mentioning text selection and/or deletion strategies. The study was also conducted in Hebrew which might differ from English. Tenhaken and Scheibner-Herzig (1988) conducted a similar study in English. In their study 64 eighth-grade students were given an oral interview and a cloze test. The researchers found that there was a high correlation between the interview scores and the cloze test scores of the students. However, this study fails to address how to specifically prepare cloze tests to better measure speaking skills as the study makes use of only one cloze test. Different cloze tests with different text selections and deletion methods (in addition to this fixed deletion ratio test) might have yielded different results. Moreover, as we have seen from the previous examples, the cloze procedure requires the integration of several language and

general skills. Therefore, as this study was conducted at a primary school level, the results may turn out to be very different with university level students with the change in age, level and cultural experience.

Statement of the Problem

The cloze technique is regarded to be a useful tool in a wide range of areas by many researchers in the field (Froehlich, 1985; Oller Jr., 1972; Stubbs & Tucker, 1974; Wipf, 1981; Yamashita, 2003). As the procedure is a quick and economical method of assessing a wide range of different language skills (Brown, 1980; Oller Jr., 1973), a lot of research has been done on how and where to use cloze tests as effectively as possible (Brière et al., 1978; Froehlich, 1985; Jacobson, 1990; Stansfield, 1980; Steinman, 2002; Stubbs & Tucker, 1974; Taylor, 1953; Torres & Roig, 2005; Wipf, 1981; Yamashita, 2003). Oller Jr. (1974) stated that:

“The ability to anticipate elements in sequence is the foundation of all language skills. Because of its naturally high redundancy, it is almost always possible in the normal use of language to partially predict what will come next in a sequence of elements.” (p. 443)

Since then, many researchers (Darnell, 1968; Hadley et al., 1997; Yamashita, 2003) have tried to see if cloze tests can assess the grammar, vocabulary, reading, and listening skills of students. However, there has not been much research done to find if cloze tests are adequate tools in measuring the productive skills (writing and speaking) of L2 learners. Although some researchers (Hanania et al., 1986; Shohamy, 1982; Tenhaken et al., 1988) have tried to look at the cloze procedure from a productive point

of view (see Chapter Two for details), their studies recommended further research with some also having different shortcomings.

Even though cloze tests have been a valuable testing tool for many teachers since their first appearance in 1953, the researcher does not think that there has been enough use of them in Turkey. This lack of interest may be attributed to a lack of local research on the topic. After over half a century since their first use, cloze tests are still not popular with Turkish researchers and have not gathered even a fraction of the interest that they have in the international research society with not even a single study on them to my knowledge. According to my own experience, this may partly be due to some lack of understanding on some Turkish teachers' part. Cloze tests are mainly regarded as a fancier form of a fill in the blank exercise rather than a powerful assessment tool. If such misunderstanding is the case, then those who do not understand cloze tests fully may also not know how to put them into use.

Research Questions

This study will address the following research questions:

1. To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' speaking skills?
2. To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' writing skills?
3. Are there important differences between cloze test preparation and scoring methods with regard to how well they assess students' speaking and writing skills?

Significance of the Study

How to best assess students' competence levels in a foreign language has always been a vital question for researchers and teachers around the world. Even though cloze tests are regarded as powerful tools in testing by many researchers, very little research has been carried out to examine the effectiveness of cloze tests in assessing productive skills in a foreign language. The present study will fill this gap by discussing several different uses of cloze procedure (in regards to text selection, deletion method, and scoring method) to assess university level students' productive skills.

At the local level, the current research study aims to inform Turkish testers in preparatory schools by exploring a possible addition to common testing methods (oral interviews, portfolios, essay writing exams). Although, given their disadvantages, I do not believe that the cloze tests can replace the common testing methods in this area (which will be discussed in the next chapter) I think cloze test have much to offer as supplementary tests. According to my own experience, in Turkey, conducting tests to assess language learners' proficiencies in speaking and writing skills is sometimes done inadequately. Especially in language classes that only gather for three or four hours a week (for example Foreign Language 101 classes or English classes in primary and high school), productive skills are sometimes completely neglected. However, such negligence is regrettable, as those two skills are vital in language acquisition. The process of testing a medium to large population of students' productive skills is a very arduous and time-consuming one and I believe this to be the main reason behind the negligence of those two skills in testing. When their advantages are taken into

consideration, I believe, with enough research and fine tuning, specific cloze tests can be designed to help in the assessment procedure of productive skills, especially in situations where the frequent use of direct methods is not economical or possible.

Conclusion

This chapter presented the background of the study, statement of the problem, the research questions, and the significance of the problem. The following chapter will look at the prior literature on the subject. The third chapter will explain the methodology of the present study. The fourth chapter will talk about the data and its analysis. The last chapter will conclude the study commenting on the analysis.

CHAPTER II: LITERATURE REVIEW

Introduction

This study investigates the possible uses of cloze tests in assessing students' productive skills (speaking and writing). The study also focuses on how to prepare and how to score such cloze tests to get the best results in regards to assessing productive skills. The selection of the text, the deletion methods and the scoring methods all play an important role to better assess certain skills by using cloze tests.

This chapter mainly focuses on past literature about cloze tests and is divided into four different parts. The first two parts focus on testing writing and speaking. The third part briefly mentions indirect testing, as the method of assessing productive skills with the use of cloze tests is mostly indirect. The last part mainly focuses on cloze tests and their different uses.

Testing Writing

Being a productive skill and having a concrete product which can be assessed in any desirable way, testing writing is usually a straightforward process. Hughes (1989) believes that the best way of testing people's writing ability is to get them to write. The testing process in writing is mainly composed of a test taker producing a piece of written work followed by the scoring of that piece of work by the testers in a predefined way. Even though the process is generally fixed in this direct approach, writing tests differ greatly in other aspects.

Brown (2004) begins his chapter on writing with a classification of written language into three main areas. Those three areas are:

1. Academic Writing: *papers, essays, articles, theses, dissertations...*
2. Job-related Writing: *messages, letters, memos, schedules, reports...*
3. Personal Writing: *e-mails, journal entries, fiction, notes, shopping lists...*

With these three main groups already classified, Brown goes on to analyze how much writing knowledge is required by learners to be able to write different types of tests. He divides writing performance into four different areas ranging from exclusively form focused to mostly context and meaning focused as: imitative, intensive, responsive, and extensive.

Weigle (2002) defines the three latter performance types (as the first is only about proper spelling of words and using letters with their correct shapes) as:

1. Reproducing already encoded information
2. Organizing information already known to the writer
3. Inventing/generating new ideas or information

In addition to selection of the right task and performance type, Hughes (1989) suggested a few ways to ensure the reliability of writing tests. He suggested testing the candidates with as many tasks as possible while giving no choice between the different tasks. He also suggested restricting candidates to achieve a more or less unified set of results by making use of pictures, notes, and even outlines. Hughes also reminds the testers to ensure that there are long enough samples of student writing to test for the desired performance type.

Researchers (Brown, 2004; Hughes, 1989; Weigle; 2002) separate the scoring rubrics for writing skills assessment into three different groups. These groups are generally defined as:

1. Holistic Scoring: The scorers assign a single score based on the overall impression of the work in question.
2. Primary Trait Scoring: Score is a single number based on the overall effectiveness of the work in achieving the task given.
3. Analytic Scoring: The score is the sum of several aspects of writing or criteria rather than a single score.

While the literature defines the different approaches to writing rubrics, researchers agree that the choice between them depends on the purposes of testing and the circumstances of scoring.

Testing Speaking

Speaking, being another productive skill, is also mostly tested in a direct manner. Brown (2004) states that even though speaking is a productive skill which can be directly and empirically observed, it is very hard to separate and isolate oral production tasks from aural skills. He also argues that one cannot know for certain that a speaking score is exclusively a measure of oral production as test-takers' comprehension of the interlocutor's prompts are invariably colored by their listening abilities.

Hughes (1989) classifies the oral functions into five main categories which are:

1. Expressing: *opinions, attitudes, wants, complaints, reasons/justifications...*
2. Narrating: *sequence of events*
3. Eliciting: *information, directions, service, clarification...*
4. Directing: *ordering, instructing, persuading, advising...*
5. Reporting: *description, comment, decision...*

On top of these groups Brown (2004) divides oral performance into five categories ranging from exclusively phonetic parroting to creative, deliberative language as: imitative, intensive, responsive, interactive and extensive.

To cater for the different needs of the different performance categories Louma (2009) made distinctions between four different types of informational talks that are generally used in speaking tests. These talk types, listed from easy to accomplish to hard, are:

1. Description
2. Instruction
3. Storytelling
4. Opinion expressing / Justification

In addition to the selection of the right task and performance type, Hughes (1989) noted that even though there are no references to the grammatical structures, use of vocabulary or pronunciation skills, this does not mean that there are no requirements with respect to those elements of oral performance in speaking tests. He continues by

dividing up the criterial levels of performance into five main categories: accuracy, appropriacy, range, flexibility, and size.

Louma (2009) tells us about the different types of rubrics that are generally used in oral assessment. These different rubrics are categorized into four different groups as:

1. Holistic Rubrics: The scorers assign a single score based on the overall impression of the oral performance in question.

2. Analytical Rubrics: The score is the sum of several aspects of the oral performance or criteria rather than a single score.

3. Numerical Rubrics: Numerical scores on a scale are given to different aspects of performance in question.

4. Rating Checklists: Yes/No questions regarding several aspects of the performance in question.

While the literature defines the different approaches to speaking rubrics, researchers agree that the choice between them depends on the purposes of testing and the circumstances of scoring.

Indirect Testing

It can be argued that (in the field of language testing) speaking and writing skills are the only skills that can possibly be directly tested as they are productive skills that give the examiner a direct output. The receptive skills of reading and listening on the other hand are not in any way directly observable as reception happens in the brain of the recipient. Language tests usually rely on indirect methods to test such skills.

Bengoa (2008) defines direct testing as a test that actually requires the candidate to demonstrate ability in the skill being sampled. He defines the indirect counterpart as a test that measures the ability or knowledge that underlies the skill that is trying to be sampled. He gives the example of driving tests worldwide and distinguishes the 'road test' and 'theory test' as direct and indirect tests in that order. Gap filling exercises (such as cloze tests) are the possibly the most used indirect testing method in language testing.

While pronunciation tests in which the students are required to read some words out loud are direct tests in the language teaching sense, asking the students to underline the word that sounds different in a group of words is an example of an indirect test of pronunciation. Banerjee (2000) states that indirect tests such as these are mostly popular in language testing situations where large numbers are being tested. He also gives the example of the dialogue completion tasks measuring speaking in contexts where the candidature is too large to allow for face-to-face speaking tests.

While indirect tests are useful and unavoidable in certain situations, one must not forget about the disadvantages of indirect tests. As indirect tests generally do not require the learner to perform the actual action that is being assessed the content validity of indirect tests are usually low. As the content seems unrelated to the topic to some extent, the face validity of indirect tests are also negatively affected. One must also consider the probability of negative washback (the effects of testing on teaching) while working with indirect tests. While tests can have a positive washback effect, indirect tests may lead the test takers to practice the indirect testing methods instead of the actual skill that is tested, as test takers generally prefer to succeed in tests.

Studies on Cloze Tests

The cloze procedure is a major type of indirect test, which is used in the testing of many language skills. Since their first use in language teaching in 1953 (Taylor, 1953), cloze tests have been the focus of much attention and research. Many researchers (Aitken, 1977; Brière et al. 1978; Froehlich, 1985; Jacobson, 1990; Oller Jr., 1973; Sasaki, 2000) have used cloze tests for a multitude of varying purposes. Even though there have been many improvements suggested by different researchers, the study of the cloze procedure is still ongoing. Different effects of different methods have also been studied extensively.

What are cloze tests

The cloze test is prepared by first selecting a text of variable length and type. The length and type of the selected passage depends on the purpose of the text. A reading comprehension test might require a text that was read previously, while a proficiency or grammar test might require novel texts that contain the items in question (Aitken, 1977). After the text is selected, several words are deleted from the text and replaced with blanks of equal length. Hadley and Naaykens (1997) state that there are at least five different types of cloze tests: The fixed-rate deletion, the selective deletion (also known as the rational cloze), the multiple-choice cloze, the cloze elide and the C-test.

The fixed-rate deletion is mainly the deletion of every *n*th word (usually fifth or seventh but may be higher in longer texts and/or lower proficiency levels). In the multiple-choice cloze the students are given a number of choices to choose from in order to fill in the gaps. In the cloze elide the deleted words are replaced with different words

that do not fit within the text. The student is tasked with both the identification of incorrect words and their replacement with the correct words. In the C-test, instead of the deletion of whole words, the latter half of every second word is deleted (the number of deleted letters are rounded down). In the selective deletion or the rational cloze method, the tester chooses the words he wishes to delete from the text in accordance to the purposes of testing. Hadley et al. (1997) believe that the goal of this method is twofold: fine tuning the level of difficulty, and targeting specific grammatical points and vocabulary items to measure.

Scoring methods

Many different scoring methods have been utilized for cloze tests. The most prominent of those methods have been the exact word scoring method and the acceptable answer scoring method. In the exact word scoring method, the tester only checks for the exact words that were originally used in the text and only gives points if those words are used. Meanwhile, in the acceptable answer scoring method, acceptable answers for every blank are determined (mostly with the help of native speakers) and points are given for every acceptable word. There are also some innovative new methods by different researchers who try to find which scoring method measures the different abilities of the students.

Clausing & Senko (1978) designed two different scoring methods as alternatives to the exact scoring method. They named their methods as ‘The Quick Performance Measurement and Feedback Technique’ and ‘The Three-Stage Scoring Hierarchy for Partial Credit’. They claimed that these two methods “facilitate a more accurate and

integrated utilization of the cloze test” (p. 74), going beyond the exact word scoring method. They did not, however, compare the results with the exact word scoring method.

In the first technique, they awarded two points for the exact word and fully acceptable answers, one point for partially acceptable answers with small mistakes, and zero points for unacceptable answers. The acceptability of each answer was judged by the teacher, hence the name ‘quick’ in the technique.

In the second technique, they used a complex system which would award partial credit for different aspects of each answer. Context (45% of total score) was deemed more important than grammar (40%) which in turn was deemed more important than orthography (15%). All three of the aspects were again divided into four groups and awarded points accordingly. Those four groups were: correct, minor mistake, major mistake(s), and completely wrong. They also divided the scoring into three steps – one for each aspect – and used an elimination method which eliminated an answer outright if it scored below minor mistake in any of the categories. They concluded that the second technique to be an arduous practice and did not lead to a significant change in the outcome.

Kobayashi (2002) looked at three different scoring methods and the effects of item characteristics on cloze test performance. He used a fixed deletion ratio of every 13th word with four different texts on two topics (eight tests in total). His three different scoring methods were the exact word scoring method, semantically and syntactically acceptable word scoring method and semantically acceptable, but syntactically

unacceptable word scoring method. The acceptable alternative answers were defined with the help of six native speakers prior to scoring. The results showed that the two acceptable-word scoring methods had higher reliability than the exact word scoring method, although the differences were not significant. The correlations between the results of all the scoring methods and the results of a proficiency test were found to be moderately high (no exact value given). The researcher also reports that the correlations of the two acceptable methods with the proficiency test were slightly higher than the exact word scoring method.

Kobayashi also divided the items he examined into different groups according to their characteristics. The main characteristics he looked at were: (a) content and function words, (b) parts of speech, (c) word frequency, (d) the number of occurrences of a word in the text, (e) alternative answers, (f) syntactic variation, (g) the amount of context, and (h) knowledge base. For content and function words, the researcher reported that the reliability scores of both groups were low (especially with content words) for the exact word scoring method. He interestingly found that the lower reliability estimates of the content word items with the exact-word scoring method rose dramatically when one of the acceptable-word scoring methods was applied. Some of the results of content word items within acceptable word methods were even higher than those of function word items. Even though these results suggest that acceptable-word scoring methods are better tools in testing comprehension than the exact-word scoring method, Kobayashi cautiously recommends further research as the statistics varied greatly depending on the types of words deleted.

Brown (1980) compared the relative merits of four different scoring methods for scoring cloze tests. He used a passage from an intermediate ESL reader with a total of 50 blanks at seven word intervals. He compared the exact-word scoring method, the acceptable-answer scoring method, the multiple-choice format and clozentropy (a format which logarithmically weights the acceptable answers according to their frequency in a native-speaker pretest and gives fractions of scores) method. The open-ended format of the text was first pre-tested on 77 native English speakers, and the answers provided the basis for acceptable words and clozentropy. The open-ended test was also pre-tested on 164 students of English with intermediate to advanced levels to provide distracters for the multiple choice format. The final products of the pre-testing period were then applied to 112 (55 open-ended, 57 multiple-choice) students.

To compare the different scoring methods, Brown designated seven criteria and scored the scoring methods on a scale from one to four according to each criterion, with one being the best and four being the least successful in that criterion. The seven criteria were: Standard error of measurement, validity, reliability, item facility, item discrimination, usability (development), and usability (scoring).

Brown's results show that there was no difference between the validity of the four tests. The scores of the tests according to the seven criteria (with validity at one for every test) were thus: the acceptable-word scoring method (12), exact-word scoring method (13), the multiple-choice format (16) and clozentropy (22).

Even though the numbers rank all four methods with a single score, which may lead to the perception of the acceptable-word scoring method being the best method while ranking clozentropy last, Brown was unsure of such absolute ranking as all tests had different strengths and weaknesses. Brown concludes that

“...regardless of the scoring method used, the cloze procedure is perhaps deceptively simple. It is easy to develop and administer, and seems to be a reliable and valid test of overall second-language proficiency... However, the final decision on which scoring method to employ must rest with the developer/user, who best knows the purpose of the test, as well as all of the other considerations involved in the particular testing situation. (p. 316)”

Oller Jr. (1972) conducted a study which examined five different scoring methods through three different difficulty levels for cloze tests of proficiency. The participants were 398 learners of English taking the University of California ESL placement exam. There were 210 graduate and 188 undergraduate students from 55 different language backgrounds. The three cloze tests were constructed from an elementary, a lower-intermediate, and an advanced level text. Each text was about 375 words and the tests were constructed by the deletion of every seventh word. Of the students, 132 took the first, 134 took the second, and 132 took the third test by random selection.

Oller Jr. used the exact-word scoring method (M1), acceptable word scoring method (M2), and three different methods of his own devising (M3, M4, and M5). To calculate the scores according to those three different methods, he differentiated five different response types. Those response types were:

“...(a) restorations of the original words, EXACT WORDS; (b) entirely acceptable fill-ins, ACCEPTABLES; (c) responses that violated long-range constraints, e.g., "the man" for "a man" when no man has been mentioned previously, L-R VIOLATIONS; (d) responses which violated short-range constraints, e.g., "I goes" for "I go," S-R VIOLATIONS; and (e) entirely incorrect fill-ins or items left blank indicating complete lack of comprehension.”
(p. 153)

Oller Jr. assumed that S-R violations were more serious than L-R violations and designed his three scoring methods accordingly. These scoring methods were:

M3: 4 (Exact-Words) + 3 (Acceptable Words) + 2 (L-R Violations) + S-R Violations

M4: 2 (Exact-Words + Acceptable Words) + 2 (L-R Violation) + S-R Violations

M5: 2 (Exact-Words + Acceptable Words) + L-R Violations + S-R Violations

The scores of M1 through M5 were correlated with the UCLA ESL Placement Examination's (and its parts of reading, vocabulary, grammar, and dictation separately) scores using Pearson correlation coefficients. The results clearly showed that the acceptable word scoring method (M2) was superior to the exact word scoring method (M1). However, Oller Jr. did not find a significant difference between methods M2 to

M5. He concluded that, since M3, M4 and, M5 were considerably more difficult to use, the acceptable word scoring method (M2) was the most preferable out of the group. Furthermore, Oller found that the acceptable-word scoring method was superior in terms of item discrimination and validating correlations regardless of the difficulty level of a test. Finally, Oller concluded, by using the partial-correlation technique with the different parts of the UCLA exam, that of the test-types investigated, cloze tests tend to correlate best with tests that require high-level integrative skills.

In conclusion, most of the research conducted around the scoring techniques found the acceptable word scoring method to be a significant improvement over the exact word scoring method. However, the attempted improvements made to the acceptable word scoring method were generally found to be less effective and unpractical.

What do cloze tests measure

As we have already seen in the few examples above, cloze tests have been found to correlate well with different proficiency tests and their specific parts. Many researchers have found that cloze tests can measure a wide range of language skills and competence. The findings, while being controversial and requiring further study when viewed on their own, suggest that cloze tests are valid tests of overall proficiency (Aitken, 1977; Darnell, 1968; Oller Jr., 1972; 1973; Stubbs et al., 1974).

Aitken (1977) wrote that cloze tests are valid and reliable second language tests. His paper mainly discusses the construction, administration, scoring and interpretation of cloze tests of overall language proficiency. He finishes his paper by concluding:

“...I have found, after having constructed, administered and scored over a thousand cloze tests to ESL students in the last three years, that the cloze procedure is an extremely simple, yet valid language proficiency test. ... Cloze tests yield more “miles per gallon” of sweat spent in test construction than most ESL teachers realize...” (p. 66)

Brière et al. (1978) conducted a large study on native English speakers' achievement in four different foreign languages: German, Japanese, Russian, and Spanish. In their study, they tried to find whether cloze tests were sensitive enough to separate the language competence of first, second and third semester students in such diverse languages. They conducted their study with a group of 204 students. The test was created from narrative texts which were all approximately 500 words in length. The deletion was done at a seven-word interval and there were 50 test items in each text. All of the texts were at second semester level and they were selected from passages that would contain no unknown vocabulary at the second semester level. The tests were scored with the exact-word scoring method. The ANOVA scores for each language were determined to be statistically significantly different for every semester at a $p < .001$ level, proving the effectiveness of cloze tests in distinguishing between students at different levels.

Stubbs and Tucker (1974) tried to find out whether cloze tests were a valid measurement of ESL proficiency. They also compared the exact word scoring method and the acceptable word scoring method. They conducted their research with the help of 211 participants who were also taking the English Entrance Examination (EEE) of the

American University of Beirut. The cloze test of 50 items was prepared from a 294 word narrative text by systematic deletion of every fifth word and every test was scored twice with the different scoring methods.

The researchers found a significant positive correlation ($r=0.76$, $p<0.01$) between the cloze test scores and EEE scores of the students. The scoring methods had a very high correlation ($r= 0.97$, $p<0.01$) between each other even though their means were different. They concluded their paper by stating "... this technique constitutes a powerful and economical measure of English- language proficiency for non-native speakers as well as a useful diagnostic tool for the classroom teacher." (p. 241)

Heilenman (1983) examined cloze tests to see if there are any correlations between the cloze procedure and foreign language (French) placement tests. The research was conducted on 15 students from Northwestern University. All the students were given the university's placement test and a cloze test. The cloze test was prepared from a 255 word narrative passage by the systematic deletion of every seventh word, resulting in a forty-five item test. The tests were scored using the acceptable word scoring method. The results suggested that cloze tests measured the same thing as the placement test, with a very high positive correlation ($\rho = .93$, $p<.01$) between the two tests. Heilenman concludes her article by recommending the use of cloze tests in placement tests unless the placement instrument is specifically in relation with the program objectives.

Darnell (1968) conducted a study using his clozentropy method to test foreign students' ability in English in comparison to their native peers. The subjects were 48 foreign and 200 native speaker students at the University of Colorado. All the students were administered two different cloze tests prepared from passages of continuous prose (one difficult and one easier). The non-native test takers were also administered a TOEFL test. The answers of the native speakers were used in a complicated scoring procedure developed by the researcher (Darnell's clozentropy).

Regression-correlation analysis showed a high correlation between the TOEFL test and cloze test scores (.838, $p < .001$). Darnell also highlighted an interestingly high correlation between the cloze test scores and the listening comprehension subtest scores of the students in TOEFL (.736). He explains that the correlations between the cloze test scores and all the TOEFL subtest scores are highly significant, but he finds this particular correlation being the highest "...doubly surprising since the CLOZENTROPY test might, on a priori grounds, be thought to have more in common with the English Structure or the Reading Comprehension subtests..." (p. 19)

Yamashita (2003) experimented to see whether cloze tests can measure text-level processing ability. He chose six skilled and six less skilled readers to fill in a 16-item gap filling text. The text was a modified passage from an EFL textbook and the 16 items were chosen to be words that would require text level judgment to fill in correctly (cohesive devices and key content words). The students were also asked to provide think-aloud protocols to help in uncovering which level of knowledge they used to

answer the questions. The results were calculated using a semantically acceptable answer method and the reliability of the test was 0.84 (Guttman split-half procedure).

The researcher found that skilled readers had higher mean scores than less-skilled readers on a consistent and significant level. He also found (with the help of the think-aloud protocols) that all of the students tended to use text-level or sentence-level information, which means that the items generally tested local or global-level reading ability. These two findings seem to support the idea that cloze tests are good measurements of reading skills, especially reading comprehension.

Hadley and Naaykens (1997) conducted a study to find whether a selective deletion cloze test correlates well with a traditional, grammar-based test. In their study 22 false beginner Japanese students were given both a cloze test and a grammar test. The cloze test was prepared with the selective deletion of 25 words from a 133 word narrative text; meanwhile the grammar test was taken from the students' textbook.

The answers to the cloze tests were scored by a native speaker using the acceptable answer scoring method. The scores of the cloze tests and the grammar test were then correlated using the Pearson's Correlation Coefficient. The results showed that selective deletion cloze test scores had a high correlation with the grammar test scores ($r = .72$). The researchers concluded that, with careful consideration, cloze tests had a high potential for reliability. They also suggested further research considering correlation studies between cloze tests and oral/aural proficiency tests.

Hanania and Shikhani (1986) looked at the interrelationships between a standard proficiency test, a writing test, and a cloze test. The standard American University of Beirut proficiency test was used together with four different cloze tests (prepared using narrative texts out of textbooks) of 50 blanks (randomly distributed, one for every student) and ten different writing topics (paired with each other and randomly distributed to the students, self selection of one of the two topics by the student, 250 words). The cloze tests were scored with the exact word scoring method while the writing scores were calculated using a holistic rubric.

All of the 1572 students that applied for the American University of Beirut were given all three tests. Using the AUB proficiency test scores, 20% of the participants (N= 337) were selected as a sample from the group and had a mean score equal to the general mean. Nearly half (159) of these students were the students who had taken the minimum score of 500 for admission to the university, and were above average.

The results showed that all three tests correlated highly with each other. The cloze procedure had positive correlations of $r = .79$, and $r = .68$ with the proficiency and writing scores respectively. The correlations were, on the overall, higher within the general sample (N= 337) compared to the restricted sample (N= 159). In conclusion, the researchers stated the effectiveness of cloze tests as supplements to tests of proficiency. However, they also recommend careful preparation, and validation before using cloze tests.

Shohamy (1982) investigated the correlation between a Hebrew Oral Interview speaking test and the cloze procedure. The participants were 100 University of Minnesota students who were (or had been) enrolled in Hebrew language classes and six native Israelis. Two cloze tests were prepared using narrative texts (one easy and one difficult) with the deletion of every sixth word. Each test had a total of 50 blanks (300 total words). The cloze tests were scored twice, using both the exact word scoring method and the acceptable answer scoring method. The oral interviews were scored on a scale of zero to five in four different oral proficiency categories of grammar, vocabulary, pronunciation, and fluency.

The results indicate a high positive correlation across all levels. The combined scores of the acceptable answer scoring method for both cloze tests correlated at $r = .856$ with the total scores of oral interviews. Even though the results were high, Shohamy does not recommend substituting the more efficient cloze procedure for the oral interview in the assessment of speaking proficiency. She points out that substituting one test for another is a complex issue with multiple factors and advises extreme caution to decision makers.

Tenhaken and Scheibner-Herzig (1988) studied the cloze procedure from an oral communicative point of view. In their study, 64 eighth-grade students were given an oral interview and a cloze test. The cloze test was constructed from a passage of 364 words by a systematic deletion of every sixth word to produce 48 test items. The tests were then scored by using the exact word scoring method, the acceptable word scoring method, and the clozentropy (frequency weighted scoring method). Twenty-one native

English speaking students of the same age were also given the cloze test to collect data on the acceptable words.

The researchers found that there was a high correlation ($r = .54, .56, \text{ and } .58$ with $p < .001$ respectively) between the interview scores and the cloze test scores of the students. They also mentioned that the three methods of scoring did not differ (with $r = .96, .96, \text{ and } .98$ with $p < .001$ for all three correlations). The researchers concluded their article by stating that "...a carefully chosen cloze test adapted to the level of learners is a suitable instrument to estimate the level of these learners in oral expression and rank them." (p. 166)

Conclusion

In this section, I mainly looked at the literature on cloze tests. The literature was mostly focused on how to use cloze tests, how to better cloze tests and how to score cloze tests. It is clear from the literature that further inspection and research into the field of cloze tests is required. Especially the lack of studies in the area of using cloze tests for testing speaking and writing skills opens up research possibilities in the field. Moreover, the different deletion methods proposed by different researchers showed varying results which requires further testing.

Even though the research into cloze tests varied greatly in terms of the deletion methods and scoring methods it is surprising that none of the researchers tried to use different text types. If actual language performance is based on the expectancy of elements in context as Oller (1974) claimed, then I believe we should give our students actual contexts in which they would use speaking/listening skills by the use of dialogues.

Therefore, I decided to use both article type texts and dialogue script type texts for my research as well as different deletion and scoring methods.

From the research that has been discussed so far, it seems like the research into cloze tests has decreased since the 80s. This decline might have been caused by the shifts in general foreign language teaching trends from the grammar translation method to the audio-lingual method, and communicative language teaching method. There has also been a shift in trends in the language testing area from indirect tests to more direct ones (Shohamy & Reves 1985). However, in spite of these changes in the language teaching and testing methods, cloze tests are still valid means of assessing a wide variety of language skills. In a recent interview (Lazaraton, 2010) Dr. Elena Shohamy pointed out to the usefulness of the cloze procedure as a multi-purpose testing tool. A tool, that is, still, albeit limited, used in language teaching institutions in Turkey. Moreover, there had been no studies (to my knowledge) in Turkey using cloze tests, which is a big gap in the literature.

In the next chapter we will talk about the methodology of the present research.

CHAPTER III: METHODOLOGY

Introduction

This study investigates the effectiveness of cloze tests in assessing the speaking and writing skill levels of Manisa Celal Bayar University Preparatory School students. The study also tries to find out if there are differences in the success levels of such tests in assessing speaking and writing skill levels when the conditions are changed. It examines the text selection, deletion methods and scoring methods to determine whether there are any differences in the success levels.

The research questions for this study are:

1. To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' speaking skills?
2. To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' writing skills?
3. Are there important differences between cloze test preparation and scoring methods with regard to how well they assess students' speaking and writing skills?

In this chapter, information about setting, participants, instruments, data collection procedure and data analysis will be provided.

Setting

This study was conducted at Manisa Celal Bayar University Preparatory School. Some of the departments in CBU require a certain degree of English Language proficiency before they accept freshmen. Every year, new students are given a proficiency test of English to check whether their proficiency levels are satisfactory. Students who fail the proficiency exam must attend a mandatory one year English preparatory school before moving on to their departments. Every year, the students that fail to pass the proficiency test are also given a placement test. The placement test scores are then used to divide the students into two or three different levels in the beginning of the year. These categories are then further divided into classroom sized chunks (20 students each). The criterion for this last division is the departments in which the students will be studying in the next year. The students are exposed to between 24 to 28 hours of English every week.

In the 2010/2011 academic year, there were three different levels of students. They were divided into groups as A (Pre-intermediate), B (Elementary), and C (Beginner). In that academic year there were three different lessons which the students had to participate in. There was a general English lesson which included grammar, vocabulary, speaking, listening and reading skills. This general English lesson used the coursebook *New English File* by Clive Oxenden, Christina Latham-Koenig, and Paul Seligson (Oxford) for 16 - 24 hours per week. There were also Academic Reading and Academic Writing lessons for four hours each. Academic Reading lessons followed *Inside Reading* by Arline Burgmeier, Lawrence J. Zwier, Bruce Rubin, and Kent

Richmond (Oxford) and Academic Writing lessons followed Effective Academic Writing by Alice Savage, Patricia Mayer, Masoud Shafiei, Rhonda Liss, and Jason Davis (Oxford) respectively.

How is Writing Taught and Assessed in CBU?

The students are often required to write up lab reports, essays, homework articles, or small annotated bibliographies during their studies in their departments in CBU. Thus, writing is considered to be an important skill by the administration and lecturers in the Department of Foreign Languages. The students start taking regular writing lessons after they are finished with beginner and elementary courses. Students that are already at lower intermediate level start practicing their writing skills immediately.

The writing lessons begin with paragraph writing exercises and lectures on how to write effectively and continue on to writing letters, e-mails, reports, stories, and essays. The writing skill is generally taught by one teacher per class throughout the year. The teacher grades and gives feedback on in-class participation. In the academic year of 2010 – 2011 the writing lesson implemented a portfolio in which everything produced in the lesson was contained.

The writing skills of the students are assessed regularly in the Foreign Languages Department. The assessment is threefold. First, there is the assessment of in-class participation and the portfolio. This is done according to the criteria chosen for that year's portfolio and the actual scoring is done by the respective instructors of each class. Second, there are writing sections in almost every quiz that is given throughout the

semesters. The writing parts in those quizzes range from 5% to an actual writing quiz of 100% of the total quiz score. Lastly, the midterms and end of semester exams feature writing parts which play a crucial role in the decision of passing or failing the student. The writing scores of the students are generally calculated with the use of an analytical rubric which is designed specifically for every different task by the testing office. These two types of scoring are generally done by two separate teachers and are checked for inter-rater reliability by the testing office using the joint probability of agreement method.

How is Speaking Taught and Assessed in CBU?

Speaking is, in my opinion, one of the weak points of the Foreign Languages Department in CBU. Even though some of the students are required to use spoken English in their classroom participation when they get to their departments, there is no special attention paid to speaking as a skill. The speaking skill is generally integrated into the General English and Academic Reading lessons of the students.

The assessment of speaking in CBU is the main reason that I have begun research on this subject. Even though the students are assessed for all other skills throughout the year the speaking skills of the students are assessed only once in two whole semesters. The proficiency exam, the placement exam, the six midterms, and the numerous quizzes do not have any speaking assessment whatsoever. The only time the students are tested in speaking is as a part of the end of the year exam which they take in late May.

This speaking exam is carried out as an interview between a student and two lecturers. The test consists of a series of questions and prompts which are prepared before the exam and given to the instructors who will proctor the speaking test. The testers are provided with a sufficient number of prompts and dialogue topics so that they do not use the same topic twice during an exam. The exam is conducted in the regular classrooms of the students with lecturers different than their classroom teachers interviewing the students. Both of the interviewers score the interviewee separately and combine their results to make up a final score. The raters are also checked for inter-rater reliability (again using the joint probability of agreement method) after they are finished with the testing. Generally a holistic rubric is used in the scoring of the students.

Participants

Sixty students from Celal Bayar University Foreign Languages Department participated in this study. The participants were the students of all the three A level courses in the department. A level students were chosen for this study because they were the only level that had had writing lessons since the beginning of the first semester, while the other two levels had begun taking writing lessons in the second semester. The age of the participants ranged from 17 to 21, with an average of 18.9. There were 14 male and 46 female participants.

All of the 60 participants were students of the English Language and Literature Department. They studied English 24 hours a week during that year. Their schedule included 16 hours of General English, four hours of Academic Writing and four hours of Academic Reading. Each class had four different instructors throughout the year, two

General English instructors, one Academic Reading instructor, and one Academic Writing instructor.

This study was conducted near the end of the second semester. By that time the participants had already completed intermediate level and were at the beginning of the upper-intermediate level instruction. The participants graduated from the prep-school at an intermediate (+) level with four weeks of extra training after they finished their intermediate level studies. All participants were willing to participate in the study and they were assured that their names and scores would be kept confidential.

Instruments

Six different cloze tests (constructed out of two different texts), a speaking interview (prepared and scored by the home institution), and a writing test (prepared and scored by the home institution) were the instruments used to collect data in this study.

The Cloze tests

Two different texts were selected by the researcher to be mutilated in three different ways to create six cloze tests. One of the texts was selected from *The Wall Street Journal* and the other was taken from the script of a movie called *The Shining*. While the text taken from the journal is an article about a businessperson, the other text is taken from an interview scene and consists of a dialogue mainly between two characters in the movie. I decided to use both article type texts and dialogue script type texts for my research as I believed that dialogue scripts would be better suited to activate background knowledge of the students about speech patterns and thus be more successful in assessing speaking skills than a narrative article. I predicted that the

dialogue type texts would have higher correlations with a speaking test while the article type texts would have higher correlations with a writing test. Both texts were simplified by the researcher before they were used. The texts were mutilated into cloze tests using three different deletion methods (see Appendix A for the un-mutilated texts).

The Article Type of Cloze Tests

The article taken from the Wall Street Journal is the story of a successful businessperson in the catering business. The article was first simplified by changing some of the low frequency words with high frequency words and then analyzed for word frequency. VocabProfiler (<http://www.lexutor.ca/vp/eng/>) was used for the analysis of the text. Out of the 461 word tokens and 198 word types in the text, 458 tokens (195 types) were in the first 2000 word range. There were only three words that were either in the academic word list or off-list (three occurrences). The VocabProfiler also helped in determining the function and content words within the text, by listing the content and function words within the text separately.

After the analysis of the text it was mutilated into three different cloze tests by:

1. The deletion of every 13th (± 4) function word.
2. The deletion of every 13th (± 4) content word.
3. The deletion of every 13th (± 1) word.

The function words were defined by the researcher to be: prepositions, articles, modal verbs, auxiliary verbs, pronouns, and conjunctions.

The content words were defined by the researcher to be: nouns, verbs, adjectives, and adverbs.

The deletions were done in this way in order to measure the students' ability to fill in function words, content words, and random selection words in an article. All tests had a few sentences in the beginning and in the end of the tests without any blanks to provide context. The deleted words were chosen as close to the 13th word mark as possible (The ± 1 range in the random selection method was only utilized, at the discretion of the researcher, where there were no context clues and thus very little chance of guessing the exact word in the text). The 13th word mark was chosen so that there was a total of 30 blanks in each test (see Appendix B for the tests).

The Dialogue Type of Cloze Tests

The dialogue taken from the script of the movie *The Shining* is a job interview. The article was first simplified by changing some of the low frequency words with high frequency words and then subjected to the VocabProfiler (<http://www.lexutor.ca/vp/eng/>) for analysis. Out of the 831 word tokens and 312 word types 817 tokens (299 types) in the first 2000 word range. There were only 13 words that were either in the academic word list or off-list (14 occurrences). VocabProfiler also helped in determining the function and content words within the text, by listing the content and function words within the text separately.

After the analysis of the text, it was mutilated into three different cloze tests by:

1. The deletion of every 22nd (± 8) function word.
2. The deletion of every 22nd (± 8) content word.
3. The deletion of every 22nd (± 1) word.

The function and content words were defined in the same way with the Article type of cloze tests.

The deletions were done in this way in order to measure the students' ability to fill in function words, content words, and random selection words in a dialogue. All tests had a few sentences in the beginning and in the end of the tests without any blanks to provide context. The deleted words were chosen as close to the 22nd word mark as possible (The ± 1 range in the random selection method was only utilized, at the discretion of the researcher, where there were no context clues and thus very little chance of guessing the exact word in the text). The 22nd word mark was chosen so that there was a total of 30 blanks in each test (see Appendix B for the tests).

Scoring The Cloze Tests

The cloze tests were given to the students in sets of two (one from each text type). The proctor made sure that none of the students were given the same deletion type test for both of the text types (if a student received function words deletion type test for one of the tests s/he would receive either a content words deletion type or a random selection deletion type test for the other test). The tests were then scored with two different scoring methods. The two different scoring methods were the exact word scoring method and acceptable answer scoring method.

The acceptable answers for the acceptable word scoring method were determined by four different native speakers who reviewed a compiled list of non-exact answers. If a non-exact answer was considered to be acceptable by two or more native speakers, then the students received full points for that answer (see Appendix C for the acceptable answers key).

The Speaking Test

The speaking test was set up as an interview of the student by two teachers. The teachers were selected from the teachers that did not teach in the students' regular classes throughout the year. The two teachers were given a holistic rubric, 30 starter topics, and 40 main topics to choose from. The teachers were seated in the normal classroom where the students had their classes during the semester while the students were called from the schoolyard by some assistant teachers one by one for each classroom.

When the student came in, the teachers began with some warm up talk to get the student ready for the interview. Then the teachers selected one unused starter prompt from the 40 they had in front of them and told it to the student. The student (who already knew what to do as they had practiced before the exam in the classroom) responded to the prompt either by asking a question, stating an opinion or relating it to something else. After a small chat, the student was asked one of the 30 discussion topics that the teachers had not used and given a couple of minutes to think about the topic.

After the preparation time, the student and the teachers talked about the topic. Sometimes the teachers asked some additional questions or guided the student through the conversation. After a total of ten minutes, the student was asked to leave the room and the interviewers took a few minutes to grade the student in light of the interview. The interviewers did not see the grade that the other interviewer gave to the student. Both scores were then checked for inter-rater reliability using the regular joint probability of agreement method for each student. In this method the two scores given by the separate testers are compared and if the difference between them is over a certain threshold the scores were regarded as not reliable. However, the testing office did not share the exact threshold value with the researcher. If the scores were found reliable the average of the two scores was considered to be the student's writing score. If inter-rater reliability was not found, then the student was called for another interview the following day.

The Writing Test

The students took the writing test as a part of their regular monthly exam. The students were given the test in their regular classroom with a time allotment of 45 minutes. The exam was in the form of an essay writing exam. The students were expected to write an essay complete with a title, introductory paragraph(s), body paragraphs and a conclusion. The exam was proctored by a teacher that did not have regular lessons with the students. After the 45 minutes of allotted time was over, the proctor collected the papers and handed them down to two raters who also did not have regular lessons with the students throughout the year. Those two raters were given

analytical rubrics by the testing office and they graded all the papers separately according to the rubric. After the grading period, the raters compared the grades that they had given to the papers according to the joint probability threshold predefined by the testing office. However, the testing office did not share the exact threshold value with the researcher. If there was inter-rater reliability the mean of both scores was taken as the writing score of the student. If inter-rater reliability did not meet the criteria then the essay was graded by two different teachers until inter-rater reliability was achieved.

Procedure

After the cloze tests were prepared, they were first piloted with a group of 40 participants at the same level as the intended participant group. Before the piloting session, the participants were given a brief explanation about the nature of the study and the nature of the piloting procedure. They were asked to provide any kind of feedback they might have during and after the piloting. During the piloting process, the students gave their insights about the cloze tests. Some of the students even commented on individual blanks. In light of the comments of those that participated in the piloting session, the cloze tests were adapted.

After the piloting sessions were over and the cloze tests were modified accordingly, the researcher began conducting the tests with the real participants. The tests were conducted in the regular classrooms of the students during regular class hours. The participants were all English Language and Literature Department students and were (by extension) English language teacher candidates. The researcher explained the aim of the cloze tests to the students and asked them for their assistance. None of the

students out of the three classes refused to participate in the research. The researcher clearly explained that these tests would not affect their grades in any way. The researcher also made it clear that incorrect answers were as important as correct answers in deducing the effectiveness of these tests in the assessment of productive skills. The participants were asked to provide their names with the answers but were assured that their names would be kept confidential and would only be used in the correlation of their scores in cloze tests with their speaking/writing scores. Some of the participants seemed very enthusiastic about the procedure and the research and wanted to hear about the results when the research was over. The researcher took the e-mail addresses of those that want to be informed about the outcome of the research to contact them later.

After the explanation about the research was finished, the researcher distributed the cloze tests so that every student received two different cloze tests (one of both text types). The researcher made sure that the students got different deletion method type tests for their two text type tests. After the distribution, the researcher once again reminded the students that the wrong answers were as important as the correct answers and asked them to complete the tests on their own. The students were reminded that every blank should be filled with one word and there might be more than one possible correct answer. The researcher told the students that they had 45 minutes to complete the cloze tests.

During the 45 minutes the researcher was present in the classroom to clarify any points that might be raised by the students about the exam. After the examination was done and the papers were collected, the students were given the un-mutilated texts for both of the test types as an answer sheet.

Data Analysis

The data collected from the cloze tests, the speaking test and writing test were subjected to standard descriptive analysis and normality tests. After the normality tests were done the data were analyzed using Pearson's Correlation Coefficient and Kendall's tau_b to see how much correlation there was between the cloze test scores and interview or writing scores of the students. The correlation analyses were done as a whole group as well as separated groups according to the scoring methods and deletion methods. The methods were compared with each other to check which scoring, text selection, and deletion method correlated best with speaking/writing scores.

Conclusion

This chapter provided detailed information about the setting, participants, instruments and materials used in the study, the procedure of data collection, and the data analysis methods. The next chapter will present the results of the data analysis, and discuss the possible implications.

CHAPTER IV: DATA ANALYSIS

Overview

This study was designed to investigate the effectiveness of cloze tests in assessing the productive skills of writing and speaking amongst university level foreign language learners of English. The study also aimed to find out whether there are important differences between the preparation and scoring methods of cloze tests with regards to how well they assess students' writing and speaking skill levels. The study was conducted at Manisa Celal Bayar University Department of Foreign Languages.

The research questions for this study are:

1. To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' speaking skills?
2. To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' writing skills?
3. Are there important differences between cloze test preparation and scoring methods with regard to how well they assess students' speaking and writing skills?

The participants were 60 (upper) intermediate level students who were studying at the English Language Preparatory School of the department at the time. Two different texts were selected and turned into cloze tests by using three different deletion methods. Two out of these six different cloze tests were then administered to 20 students each.

The students were also tested for their speaking and writing ability as a part of their regular curriculum.

This chapter will present an analysis of the data provided from the cloze tests, writing test, and the speaking test.

Data Analysis Procedure

The study is mainly a quantitative one. The data were gathered through several tests which were administered by the researcher and the students' regular teachers. Data collected from the test results were analyzed using Statistical Package for the Social Sciences (SPSS) version 16.0.0 by IBM.

The cloze tests were first scored using the exact word scoring method by the researcher. Afterwards, the answers of the students were examined by four different native speakers. Any non-exact answer that was accepted by two different native speakers were determined to be acceptable answers. The researcher scored the cloze tests for a second time using the acceptable answer scoring method.

The writing exam was in the form of essay writing. The essays were scored by two different raters separately using an analytical rubric. The raters did not know the name of the student they scored and did not know the score given to the same paper by the other rater. The mean of the two scores was then taken to provide the final writing score.

The speaking exam was in the form of an interview between the test taker and two raters. The raters used a holistic rubric to score the performance of the test takers. The raters did not know the test takers personally (did not meet in the classroom before)

and did not know the score given to the student by the other rater. The mean of the two scores was taken to provide the final speaking score.

Unfortunately, it was not possible for the researcher to get the reliability statistics from the testing office, which only agreed to disclose the scores of the students.

The Results of the Six Cloze Tests

Six different cloze tests were designed to find whether text selection and deletion methods had any effect on the ability of the tests to assess speaking and writing skill levels. Furthermore, two different scoring methods were used in the scoring of all six tests to check whether different scoring methods have any effects on the effectiveness of the cloze tests in the assessment of productive skill levels. After the scoring, a descriptive analysis was done on all the twelve different types of data. The results can be seen in Table 1.

Table 1 – Descriptive Statistics for the Cloze Tests

CLOZE TESTS			DESCRIPTIVES		
Text Type	Deletion Method	Scoring Method	N	Mean (/30)	SD
Article	Content Words	Exact Word	20	10.55	3.76
		Acceptable Answer	20	13.70	4.02
	Function Words	Exact Word	20	18.05	2.72
		Acceptable Answer	20	20.05	3.44
	Random Selection	Exact Word	20	14.05	3.69
		Acceptable Answer	20	17.20	4.44
Dialogue	Content Words	Exact Word	20	7.95	2.46
		Acceptable Answer	20	12.65	3.61
	Function Words	Exact Word	20	12.15	5.77
		Acceptable Answer	20	13.90	6.23
	Random Selection	Exact Word	20	8.15	2.39
		Acceptable Answer	20	12.55	2.92

According to the descriptive analysis of the cloze test results, the acceptable answer scoring method provided higher means (+1.75 to +4.70) than the exact word scoring method. The students were more successful in the cloze tests that were created from an article than those that were created from a dialogue. The students were more successful when they were required to fill in function words than when they were required to fill in content words while random selection of test items was in the middle.

After the descriptive analysis, the cloze test data were tested for normality of distribution using both of the normality tests offered by SPSS: Kolmogorov-Smirnov test and Shapiro-Wilk test. These tests of normality are important as they will later determine our decisions on whether using parametric or non-parametric tests in our analysis (Dörnyei, 2007). The results can be seen in Table 2.

Table 2 – Normality Tests for Cloze Test Results

CLOZE TESTS			Significance	
Text Type	Deletion Method	Scoring Method	Kolmogorov-Smirnov	Shapiro-Wilk
Article	Content Words	Exact Word	.200	.435
		Acceptable Answer	.103	.346
	Function Words	Exact Word	.200	.810
		Acceptable Answer	.200	.565
	Random Selection	Exact Word	.062	.126
		Acceptable Answer	.011	.083
Dialogue	Content Words	Exact Word	.200	.192
		Acceptable Answer	.126	.448
	Function Words	Exact Word	.200	.176
		Acceptable Answer	.200	.048
	Random Selection	Exact Word	.200	.705
		Acceptable Answer	.200	.396

The tests for distribution showed that the data is mostly normally distributed. However there are two instances where the data seems to be not normal in distribution. The acceptable answer scoring method data for the random selection protocol of the article text type test and the acceptable answer scoring method data for the deletion of function words protocol of the dialogue text type test were found to be non-parametric in distribution.

To assess and evaluate the productive skill levels of the students, the institution designed a writing and a speaking test. The writing test was an essay writing test and it was scored by two different teachers using an analytical rubric. The mean of the two scores was used as the final score of the writing test. The speaking test was an interview with two different interviewers. Each interviewer used a holistic rubric to score the interviewee. The mean of the two scores was then used as the final score of the speaking test. After all the scoring was completed a descriptive analysis was conducted on both sets of data. The results can be seen in Table 3.

Table 3 – Descriptive Statistics for the Writing and Speaking Tests

Tests	Number	Mean	Standard Deviation
Writing	60	66.95 (/100)	19.56
Speaking	60	15.23 (/20)	4.08

After the descriptive analysis, the data were tested for normality of distribution using both Kolmogorov-Smirnov and Shapiro-Wilk tests. The results can be seen in Table 4.

Table 4 – Normality Tests for Writing and Speaking Tests

Tests	Number	Significance	
		Kolmogorov-Smirnov	Shapiro-Wilk
Writing	60	.016	.003
Speaking	60	.017	.000

Both normality tests showed that the speaking and writing tests were not normally distributed. In addition to the general usage of the data as a whole, the data will also be used in correlation with different groups according to the text type and deletion methods. Thus the data were tested for normality again after they are divided into groups. The results can be seen in Table 5.

Table 5 – Normality Tests for Writing and Speaking Tests (Groups)

		TESTS		SIGNIFICANCEe	
Skill	Deletion	Text	N	Kolmogorov-	Shapiro-
Tested	Method	Type		Smirnov	Wilk
Writing	Content	Article	20	.200	.243
	Words	Dialogue	20	.034	.009
	Function	Article	20	.189	.126
	Words	Dialogue	20	.085	.148
	Random	Article	20	.001	.001
	Selection	Dialogue	20	.082	.071
Speaking	Content	Article	20	.200	.244
	Words	Dialogue	20	.200	.028
	Function	Article	20	.080	.009
	Words	Dialogue	20	.090	.013
	Random	Article	20	.090	.005
	Selection	Dialogue	20	.178	.074

After the writing and the speaking data were separated into groups, the normality tests showed mixed results of distribution. Six out of the twelve conditions were normally distributed while the remaining six were not normally distributed.

Decision on Normality

After the analysis of distribution is done for all the data it was clear that some of the data were non-parametric while others were parametric. Further analysis could either be done using both tests and switching to parametric tests where possible, or using non-parametric tests for all further analysis. As I would like to compare and contrast the results of my correlations, I chose the second option and decided to use the same kind of testing for all further analysis. Moreover, as non-parametric tests are more conservative, using them would give more confidence than using parametric tests (Dörnyei, 2007). After the analysis of the data is finished, another table with possible parametric tests will be included and analyzed to inform the reader.

Correlations of Cloze Tests with Writing and Speaking Tests

After all the descriptive statistics and the distribution types were calculated, the data were ready to be analyzed to check for any possible correlations between the cloze tests, writing tests, and speaking tests. As I had decided to use non-parametric tests for most of the analyses, I selected 'Kendall's tau_b' for the correlation tests, which has more reliable and interpretable confidence intervals than its counterparts (Kendall and Gibbons, 1990). As I have predicted that there would be a positive correlation between cloze tests and writing/speaking tests I used one tailed tests of significance. Even though I predicted that the article type of text tests would better correlate with writing tests and the dialogue type of text tests with speaking tests, I decided to test both options for correlations between writing and speaking tests to see if (and how far) my predictions were correct.

General Correlation of the Writing/Speaking Tests with the Whole Group's Cloze Test
Scores

A correlation that encompasses all the 60 students as a whole was carried out, before splitting the test types into groups according to their deletion methods. The results can be seen in Table 6.

Table 6 – Writing and Speaking Tests Correlated with Cloze Scores (Groups Together)

Cloze Tests (Scoring Methods)	Writing Test Correlations (Significance 1-tailed)	Speaking Test Correlations (Significance 1-tailed)
Article (Exact Word)	.377* (.011)	.277* (.002)
Article (Acceptable Answer)	.487* (.000)	.287* (.001)
Dialogue (Exact Word)	.208* (.012)	.303* (.001)
Dialogue (Acceptable Answer)	.323* (.000)	.458* (.000)

Number of the students is 60. Kendall's tau_b (τ) was used to calculate correlations. *

indicates significance.

As Table 6 shows, there was an overall positive correlation between cloze test scores and writing/speaking test scores of the students. The article text type of cloze tests correlated higher with the writing test than the dialogue text type of cloze tests, while the speaking test correlated higher with the dialogue text type of cloze tests than the article text type of cloze tests. The acceptable scoring method was found to lead to higher correlations in all the four cases.

These results mean, without taking the deletion methods into consideration, acceptable answer scores in the article text type of tests had the highest correlation with the writing tests $\tau = .487$ ($p < .001$). The speaking tests on the other hand had the highest correlation with the acceptable answer scores $\tau = .458$ ($p < .001$).

Correlation of the Writing/Speaking Tests with the Different Groups of Cloze Tests

After the correlations done on all the 60 students as a whole group, it was time to group the students according to their cloze test types (the deletion methods) and see if there are differences in correlations when the deletion methods change. We are going to discuss the results of the article text type correlations first and dialogue text type correlations second.

Correlations within Groups – Article Text Type

The 60 students were first grouped by the deletion methods of their article text type of cloze tests. After the grouping, the correlations between the cloze test scores and writing/speaking scores of the students were calculated with regard to deletion method and scoring method. The results can be seen in Table 7.

Table 7 – Writing and Speaking Tests Correlated with Cloze Scores (For groups of Article Type Text)

Article Type Text Groups		Writing	Speaking
(Cloze Test Types)		(Significance 1-tailed)	(Significance 1-tailed)
Deletion Method	Scoring Methods		
Deletion of Content Words	Exact	.380*	.182
	Word	(.013)	(.145)
	Acceptable	.431*	.124
	Answer	(.005)	(.234)
Deletion of Function Words	Exact	.563*	.576*
	Word	(.000)	(.000)
	Acceptable	.689*	.477*
	Answer	(.000)	(.003)
Random Deletion Method	Exact	.465*	.085
	Word	(.003)	(.310)
	Acceptable	.537*	.143
	Answer	(.001)	(.204)

Number of the students is 20. Kendall's tau_b (τ) was used to calculate correlations. *

indicates significance.

There was an overall positive correlation between the cloze test scores and writing test scores of the students across all deletion types. However, the correlations with the speaking scores were not statistically significant except in the deletion of function words protocol. The acceptable answer scoring method was found to lead to higher correlations than the exact word scoring method in all the correlations done with writing tests.

The highest correlation with the writing scores was found to be with the acceptable answer scores in the deletion of function words protocol scores $\tau = .689$ ($p < .001$). Meanwhile, the highest correlation with the speaking scores was observed with the exact word scores in the deletion of function words protocol scores $\tau = .576$ ($p < .001$).

Correlations within Groups – Dialogue Text Type

After the article text type group correlations were done, the participants were regrouped according to their dialogue text type test groups. After the grouping the correlations between the cloze test scores and writing/speaking scores of the students are calculated in regards to deletion method and scoring method. The results can be seen in Table 8.

Table 8 – Writing and Speaking Tests Correlated with Cloze Scores (For groups of Dialogue Type Text)

Dialogue Type Text Groups		Writing	Speaking
(Cloze Test Types)		(Significance 1-tailed)	(Significance 1-tailed)
Deletion	Scoring		
Method	Methods		
Deletion of Content Words	Exact	.238	.436*
	Word	(.082)	(.006)
	Acceptable	.356*	.611*
	Answer	(.018)	(.000)
Deletion of Function Words	Exact	.521*	.420*
	Word	(.001)	(.007)
	Acceptable	.532*	.499*
	Answer	(.001)	(.002)
Random Deletion Method	Exact	.229	.133
	Word	(.088)	(.222)
	Acceptable	.263	.373*
	Answer	(.063)	(.017)

Number of the students is 20. Kendall's tau_b (τ) was used to calculate correlations *

indicates significance.

Out of the four possible correlations between the productive test scores and the random deletion method scores in the dialogue type text group, only one correlation was found to be statistically significant (the acceptable answer scores with the random deletion method correlated with the speaking test $\tau = .373$ ($p = .017$)). The correlation between the writing test scores and the exact word scores with the deletion of content words was also found to be insignificant statistically.

The highest correlation values with the speaking test were observed in the acceptable answer scores with the deletion of content words $\tau = .611$ ($p < .001$). The writing test scores were found to correlate highest with the acceptable answer scores with the deletion of function words $\tau = .532$ ($p = .001$).

Standalone Parametric Correlations Where Possible

After the correlation of the data according to the groups is finished, I also wanted to look at the parametric correlations of the normally distributed data out of scientific curiosity and to get a sense of completeness. There were a total of 11 correlations that can be determined using parametric correlation tests. I decided to use Pearson's Correlation (r) Coefficient for the parametric tests as it is the standard test of correlation for normally distributed data (Dörnyei, 2007). The results can be seen in Table 9.

*Table 9 – Parametric Correlations within Possible Groups: Writing and Speaking Tests
Correlated with Cloze Scores in groups according to deletion methods.*

CLOZE TEST GROUPS			CORRELATIONS	
Text Type	Deletion Method	Scoring Method	Writing Test (Significance 1-t.)	Speaking Test (Significance 1-t.)
Article	Content Words	Exact	.546*	.187
		Word	(.006)	(.214)
	Words	Acceptable	.674*	.179
		Answer	(.001)	(.226)
	Function Words	Exact	.667*	N/A
		Word	(.001)	N/A
Function Words	Acceptable	.794*	N/A	
	Answer	(.000)	N/A	
Dialogue	Function Words	Exact	.682*	N/A
		Word	(.000)	N/A
	Random Selection	Exact	.336	.168
		Word	(.074)	(.240)
	Selection	Acceptable	.346	.446*
		Answer	(.067)	(.024)

Number of the students is 20. Pearson's Correlation Coefficient (r) was used to calculate correlations. N/A: Non parametric data. * indicates significance.

Within the possible parametric correlation tests, the only significant correlation with the speaking test scores was found with the acceptable answer scores in the dialogue text type of tests with the random deletion method $r = .446$ ($p = .024$). The parametric correlations with the writing scores, on the other hand, produced some even higher correlations with their non parametric counterparts. The highest parametric correlation with the cloze test scores and writing scores was observed with the acceptable answer scores in the article text type of tests with the deletion of function words $r = .794$ ($p < .001$).

Conclusion

This chapter reported the analysis of the qualitative data gathered through the use of the instruments as described in the previous methodology chapter. According to the data analysis cloze tests were found to be able to measure speaking and writing skills to a certain extent. The scoring method, text type selection, and the deletion method were also found to affect the success of cloze tests in measuring productive skills. The following chapter will answer the research questions, discuss the findings, present implications, consider the limitations and provide advice on further research topics.

CHAPTER V: CONCLUSION

Introduction

This study investigated the effectiveness of cloze tests in assessing the speaking and writing skill levels of Manisa Celal Bayar University Preparatory School students. The study also tried to find out whether there are differences in the success levels of such tests in assessing speaking and writing skill levels when the conditions are changed. It examined different text selection, deletion methods and scoring methods to determine whether there are any differences in the success levels.

The study was conducted in Celal Bayar University Preparatory School with 60 intermediate (+) students from three different classes. The students were tested using six different cloze tests. The cloze tests differed in text selection and deletion methods. These cloze tests were then scored using two different scoring methods. The scores were correlated with the speaking and writing scores of the students to check whether there is a correlation between the cloze test results and the students' writing/speaking scores. The correlations were also compared to each other to find if there are differences in the success levels of the tests' ability to assess the different skills.

Correlation tests of Pearson's Correlation Coefficient and Kendall's tau_b were used to check for the different correlations between the cloze tests and speaking/writing tests. The results were given in the data analysis chapter.

In this chapter, answers to the research questions and discussion, limitations of the study, pedagogical implications of the study and suggestions for further research are provided.

Results and Discussion

This section will answer the research questions and discuss the findings.

Research Question 1: To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' speaking skills?

This question was answered by looking at the correlations between the cloze test scores and speaking scores of the students.

Cloze Test Results' Correlated as a Single Group

Each of the 60 participants was given two different cloze tests. One of those cloze tests was an article text type cloze test while the other was a dialogue text type cloze test. Both cloze tests were scored using two different scoring methods: the exact word scoring method and the acceptable answer scoring method. As a result, every student had four different cloze test scores. An initial analysis was carried out on the whole group to find out how effective the cloze tests were in determining the students' speaking skills regardless of deletion method. As the speaking scores were not normally distributed, Kendall's tau_b was used to calculate the correlations.

The dialogue text type cloze tests were found to lead to higher correlations with the speaking tests than the article text type cloze tests. The exact word scoring method scores showed a moderate positive correlation with the speaking scores $\tau = .303$ ($p = .001$). Meanwhile the acceptable answer scoring method scores showed a moderate to high positive correlation with the speaking scores as a whole $\tau = .458$ ($p < .001$).

These results, while showing a positive correlation, also suggest that dialogue text type cloze tests correlate higher with speaking scores than article text type cloze tests regardless of the scoring or deletion methods. The moderate correlation of $\tau = .458$ also seems promising at this point as the deletion method groups are not factored in yet.

Cloze Test Results Correlated Contrastively between Deletion Method Groups

Both text types of cloze tests were further divided into three groups according to their deletion methods. Six different cloze tests were created to be administered to 20 students each by using the three different deletion methods: deletion of function words, deletion of content words, and random deletion. The speaking scores of the participants were also grouped according to their deletion methods in order to compare the correlations' of the deletion methods with the respective speaking scores of the participants. Even though there were some normally distributed scores which would enable the use of parametric tests, the researcher opted to use non-parametric tests at this point to better compare the different deletion methods.

Three of the four groups' correlations done with the random deletion method and the speaking scores failed to yield significant results. The only group that yielded significant results out of the random deletion method was the dialogue text type random deletion scores calculated using acceptable answer scoring method which showed a moderate positive correlation $\tau = .373$ ($p = .017$).

While the deletion of function words correlated well with the speaking tests on all four occasions (I- $\tau = .576$ ($p < .001$) II- $\tau = .477$ ($p = .003$) III- $\tau = .420$ ($p = .007$) IV- $\tau = .499$ ($p = .002$) respectively), the deletion of content words was found to correlate higher with the speaking scores. The largest correlation between the deletion method groups and the speaking scores was observed in the dialogue text type deletion of content words condition scores calculated using the acceptable answer scoring method. There was a large positive correlation between this group's scores and the speaking scores $\tau = .611$ ($p < .001$).

The large correlation of $\tau = .611$ might indicate that the deletion method of deleting content words in a dialogue text type leads to a better assessment of speaking skills than other deletion methods. The acceptable answer scoring method was seen to give better correlations (I - $\tau = .611$ ($p < .001$) II - $\tau = .499$ ($p = .002$)) than the exact word scoring method (I - $\tau = .436$ ($p = .006$) II - $\tau = .420$ ($p = .007$)).

This might indicate that the acceptable answer scoring method is more suitable while trying to assess speaking skill levels of students. Moreover, the use of Kendall's tau_b suggests that the correlation results might have been a little bit higher were the scores normally distributed.

To answer the research question, I would recommend either a replication or further study on the subject. While the correlation of $\tau = .611$ is large, it is not large enough to be considered a replacement or a supplementary process for testing speaking. This correlation shows similar results as the prior studies discussed in chapter two and falls between Shohamy (1982) Tenhaken and Scheibner-Herzig (1988) in regards to testing speaking skills successfully. Further research which might focus on the deletion and text type suggested by this study would certainly help in determining whether cloze tests can really assess speaking proficiency. Combined with the other disadvantages of using indirect tests (see Chapter Two), these findings definitely need more work to validate them.

Research Question 2: To what extent are cloze tests a reliable and valid means of assessing preparatory school EFL learners' writing skills?

This question was answered by looking at the correlations between the cloze test scores and writing scores of the students.

Cloze Test Results' Correlated as a Single Group

Each one of the 60 participants was given two different cloze tests. One of those cloze tests was an article text type cloze test while the other was a dialogue text type cloze test. Both cloze tests were scored using two different scoring methods: the exact word scoring method and the acceptable answer scoring method. As a result, every student had four different cloze test scores. An initial analysis was carried out on the whole group to find out how effective the cloze tests were in determining the students'

writing skills regardless of deletion method. As the writing scores were not normally distributed, Kendall's tau_b was used to calculate the correlations.

The article text type cloze tests were found to lead to higher correlations with the writing tests than the dialogue text type cloze tests. The exact word scoring method scores showed a moderate positive correlation with the writing scores $\tau = .377$ ($p = .011$). Meanwhile, the acceptable answer scoring method scores showed a moderate to high positive correlation with the writing scores as a whole $\tau = .487$ ($p < .001$).

These results, while showing a positive correlation, also suggest that article text type cloze tests correlate higher with writing scores than dialogue text type cloze tests regardless of the scoring or deletion methods. The moderate correlation of $\tau = .487$ also seems promising at this point as the deletion method groups are not factored in yet.

Cloze Test Results Correlated Contrastively between Deletion Method Groups

Both text types of cloze tests were further divided into three groups according to their deletion methods. Six different cloze tests were created to be administered to 20 students each by using the three different deletion methods of: deletion of function words, deletion of content words, and random deletion. The writing scores of the participants were also grouped according to their deletion methods in order to compare the correlations' of the deletion methods with the respective writing scores of the participants. Even though there were some normally distributed scores which would enable the use of parametric tests, the researcher opted to use non-parametric tests at this point to better compare the different deletion methods.

In all the six different test types the acceptable word scoring method scores were found to have a higher correlation with the writing scores than the exact word scoring method scores. Likewise, the article text type cloze test scores had higher correlation than the dialogue text type cloze test scores in all the six different test types. All the correlations made between the article text type cloze test scores and writing scores were found to be statistically significant. The highest correlation between the cloze test results and writing scores were the large positive correlations which were observed in the article text type with the deletion of function words procedures. The correlations were $\tau = .563$ ($p < .001$) for the exact word scoring method, and $\tau = .689$ ($p < .001$) for the acceptable answer scoring method. The deletion of function words procedure in the dialogue text type group also produced large positive correlations with the writing scores. Those correlations were $\tau = .521$ ($p = .001$) and $\tau = .532$ ($p = .001$) for the exact word scoring method, and the acceptable answer scoring method respectively.

The large correlation of $\tau = .689$ might indicate that the deletion method of deleting function words in an article text type leads to a better assessment of writing skills than other deletion methods. The acceptable answer scoring method was seen to give better correlations than the exact word scoring method. This might indicate that the acceptable answer scoring method is more suitable while trying to assess writing skill levels of students.

Moreover, the use of Kendall's tau_b suggests that the correlation results might have been a little bit higher if parametric tests were employed in the correlation. As the data for the article text type with the deletion of function words procedures were found to be normally distributed, parametric correlation tests were also employed to see if higher results were available. Pearson's Correlation Coefficient (r) was used to test for the correlations of both exact word scoring method and acceptable answer scoring method scores of this group. Parametric tests indicated an even higher correlation: $r = .667$ ($p = .001$) for the exact word scoring method and $r = .794$ ($p < .001$) for the acceptable answer scoring method.

Thus the research question about the extent of cloze tests' capabilities in assessing the writing skills of the students can be answered. The standalone correlation of $r = .794$ shows that cloze tests might be effectively used in assessing the writing skills of the students to a large extent where more direct tests are unavailable or unfeasible. Moreover this correlation, is almost the same with the correlation found by Hanania and Shikhani in 1986. However, as this correlation suggests a special text type and deletion method, and is singled out from a number of different correlations the researcher suggests further study or a replication to confirm the findings in this thesis. Moreover, the researcher recommends great care in the preparation process of the cloze tests in regards to text selection, deletion methods, and scoring methods if such tests are to be used in assessing writing skills.

Research Question 3: Are there important differences between cloze test preparation and scoring methods with regard to how well they assess students' speaking and writing skills?

When the correlations were examined in accordance with the preparation and scoring methods of the cloze tests, certain patterns tend to emerge. Such patterns suggest that there are important differences between those variables with regard to how well these cloze tests assess the students' productive skills.

In regards to text selection, article text type cloze tests seem to be better at assessing students' writing skills while dialogue text type cloze tests seem to be better at assessing the speaking skills. This might be because an article is a piece of written work while a dialogue script is ultimately a piece of conversation.

In regards to deletion methods, the tests that were mutilated using the deletion of function words protocol were more successful in assessing the writing skills of the students. The deletion of content words protocol tests on the other hand seemed to be more successful in assessing the speaking skills of the students. The random deletion method generally seemed to be somewhere in between the two different methods but this deletion was unreliable at times, leading to different results. This might have occurred as the redundancy is decreased in speaking situations to give more focus on content. Some of the students in the pilot study reported that they were trying to put themselves into the speaker's shoes while answering the questions in the dialogue type cloze tests, which might indicate that while they might prefer using different (and maybe wrong) functions they would still be able to get the message across with the content.

Of the two scoring methods used, the acceptable answer scoring method was found to have better correlations with both speaking and writing skills than the exact word scoring method on all occasions. This does not really surprise me as the language is not a narrow street with only one way to convey the meaning but rather a cobweb with a high number of possibilities in context. Thus it is sometimes very hard to guess the original word even for native speakers.

To answer the research question in short, the preparation methods and the scoring method seems to be greatly influential with regard to how well the cloze tests measure speaking and writing skills.

Limitations of the Study

There are a few limitations to the study which I would like to point out. As Celal Bayar Foreign Languages Department carries out only one speaking test for the whole year, the study was based on only one set of cloze tests, and one writing test done in the same time period within two weeks of the speaking test. In addition, the study was conducted only at the level of intermediate (+) as the sole speaking test in the institution is done at this level. More speaking tests at different proficiency levels would enable further testing of cloze tests' abilities in assessing productive skills, which in turn would lead to more generalizable results.

While the total number of the participants is 60, the number of participants in each group was 20. Even though there were more students at the same proficiency level in the institution at the time of research, those other students had not been taking regular writing classes and thus were not at the same level as the 60 participants in regards to writing skills. A larger number of participants would lead to more generalizable results.

As the speaking exam (and the whole research) was conducted at the end of the semester the researcher was not able to inform the participants about the results and get their opinions about them. A few quick follow up interviews or short open-ended questionnaires would have provided more insight.

Pedagogical Implications of the Study

The results of this study indicate that cloze tests could be used to help in the assessment of productive skills to some extent. The results show that cloze tests can be used to supplement and/or reinforce some institutions' tests for productive skills' proficiency by providing additional insights about students' proficiency levels, where further (or any) direct tests are not practical or available. Considering their ease of use, preparation, and scoring in relation to their more direct counterparts in the assessment of productive skills, I would suggest that cloze tests can be a welcome addition to the direct tests that are already in place for any institution. The institution that the research was conducted in, for example, might benefit from a few cloze tests in addition to its single speaking test in the assessment of speaking skills.

Some might argue that the correlation between cloze tests and productive tests might not indicate that the cloze tests are measuring productive skills, but rather that both productive tests and cloze tests are related to a third, underlying variable, namely general language proficiency. As many researchers have pointed out that cloze tests are very reliable tests of language proficiency (Aitken, 1977; Darnell, 1968; Oller Jr., 1972; 1973; Stubbs et al., 1974) that claim might be true. However, even if this is the case, cloze tests can make a useful supplement to productive tests because we would, in general, expect students with high language proficiency to do better in productive skills than those with lower proficiency. Therefore, it seems possible to use cloze tests as reliability measures for tests of writing and speaking. With many different possible outside effects in direct tests of productive skills, such the effect of proctors or impartiality (or lack thereof) of scorers, I believe cloze test scores would be a welcome addition to the reliability measures of testing offices.

Even though some of the correlations found were large, they were not big enough to say that cloze tests can be an alternative testing method on their own. Nonetheless, the differences between the correlations of different preparation and scoring methods suggest that there might be some chance to further refine the cloze tests to make them more appropriate as alternative testing methods.

One thing that must be kept in mind, however, is the possible problems of using indirect tests like these cloze tests. As the validity (especially face validity) numbers of an indirect test may not be as high as a well prepared direct test, negative washback would also be an obstacle in the way of indirect testing as it is natural for students to focus on the objectives of the tests rather than the objectives of their learning.

Suggestions for Further Research

In regards to the limitations of this study, further research might be conducted on the topic at hand. First and foremost, a replication of the study can be suggested to confirm the several findings about the deletion methods and text selection. With more resources and time, another study with more participants, more sets of tests, or with participants at different levels of English proficiency would provide more insights. Furthermore, a larger study might be conducted to include other deletion methods, text types or scoring methods.

As the research is mainly on indirect testing, a study examining the possible washback effect that the extensive use of cloze tests might create in testing speaking/writing might be considered. Another follow up study might include surveys which might inform students, teachers, testing offices, and administrators about the results of this study and ask for their opinions on the matter.

Conclusion

This study investigated the effectiveness of cloze tests in assessing the speaking/writing skills of university EFL learners. The study also investigated whether different preparation or scoring methods cause any differences in the effectiveness of cloze tests in the assessment of productive skills. The results indicated that cloze tests are capable of assessing the speaking and writing skills of EFL learners to some extent. The results also suggest that different preparation and scoring methods of cloze tests lead to different strengths in the assessment of productive skills.

The results pointed out a possible use of cloze tests in testing speaking/writing. Although strong correlation was found between writing/speaking tests and cloze tests, the numbers indicate that the cloze tests might not be adequate measuring tools for the productive skills on their own. This suggests a supportive use of cloze tests. The differing correlations also suggested a careful use of such tests. In conclusion, cloze tests can be valuable tools, if used correctly, to help assess speaking and writing skills.

REFERENCES

- Abraham, R. G., & Chapelle, C. A. (1992). The meaning of cloze test scores: An item difficulty perspective. *The Modern Language Journal*, 76(4), 468-479.
- Aitken, K. G. (1977). Using cloze procedure as an overall language proficiency test. *Tesol Quarterly*, 11(1), 59-67.
- Banerjee, J. (2000, May 31). Direct/indirect testing. *Routledge Encyclopedia of Language Teaching and Learning*. Retrieved July 4, 2011, from <http://www.bookrags.com/tandf/directindirect-testing-tf/>.
- Bengoa, M. (2008). Approaches to testing: Direct and indirect testing. Retrieved July 4, 2011, from <http://miguelbengoa.com/elt/?p=696>.
- Brière, E. J., Clausing, G., Senko, D., & Purcell, E. (1978). A look at cloze testing across languages and levels. *The Modern Language Journal*, 62(1), 23-26.
- Brown, J. D. (1980). Relative merits of four methods for scoring cloze tests. *The Modern Language Journal*, 64(3), 311-317.
- Brown, J. D. (2004). *Language assessment: Principles and classroom practice*. London: Longman.
- Chatel, R. G. (2001). Diagnostic and instructional uses of the cloze procedure. *The New England Reading Association Journal*, 37(1), 6.
- Clausing, G., & Senko, D. (1978). On the scoring of cloze tests. *Die Unterrichtspraxis*, 11(2), 74-81.
- Darnell, D. K. (1968, October). The development of an English language proficiency test of foreign students, using a clozentropy procedure. Final Report. Retrieved from <http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED024039>.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Froehlich, J. (1985). Improving aural comprehension and spelling through a modified cloze procedure using songs and poems. *Die Unterrichtspraxis*, 18(1), 49-54.

- Hadley, G. S., & Naaykens, J. E. (1997). An investigation of the selective deletion cloze test as a valid measure of grammar based proficiency in second language learning. *Niigata studies in foreign languages and cultures*, 3(12), 111-118.
- Hanania, E., & Shikhani, M. (1986). Interrelationships among three tests of language proficiency: Standardized esl, cloze, and writing. *Tesol Quarterly*, 20(1), 97-109.
- Harris, D. P. (1985). Some forerunners of cloze procedure. *The Modern Language Journal*, 69(4), 367-376.
- Heilenman, L. K. (1983). The use of a cloze procedure in foreign language placement. *The Modern Language Journal*, 67(2), 121-126.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Jacobson, M. J. (1990). Group vs. individual completion of a cloze passage. *Journal of Reading*, 33(4), 244.
- Kendall, M., & Gibbons, J. D. (1990). *Rank correlation methods*. Oxford: Edward Arnold.
- Kobayashi, M. (2002). Cloze tests revisited: Exploring item characteristics with special attention to scoring methods. *The Modern Language Journal*, 86(4), 571.
- Lazaraton, A. (2010). From Cloze to Consequences and Beyond: An Interview With Elana Shohamy. *Language Assessment Quarterly*, 7(3), 255-279. Routledge.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Oller Jr., J. W. (1972). Scoring methods and difficulty levels for cloze tests of proficiency in English as a second language. *The Modern Language Journal*, 56(3), 151-158.
- Oller Jr., J. W. (1973). Cloze tests of second language proficiency and what they measure. *Language Learning*, 23(1), 105-111.
- Oller Jr., J. W. (1974). Expectancy for successive elements: Key ingredient to language use. *Foreign Language Annals*, 7(4), 443-452.
- Oller Jr., J. W. (1975). Research with cloze procedure in measuring the proficiency of non-native speakers of English: an annotated bibliography. *Series of Languages and Linguistics*, (13), 32.

- Robinson, W. P. (1965). Cloze procedure as a technique for the investigation of social class differences in language usage. *Language & Speech*, 8(1), 42-55.
- Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A multiple data source approach. *Language Testing*, 17(1), 85-114.
- Shohamy, E. (1982). Predicting speaking proficiency from cloze tests: Theoretical and practical considerations for tests substitution. *Applied Linguistics*, 3(2), 161-171.
- Shohamy, E., & Reves, T. (1985). Authentic language tests: where from and where to? . *Language Testing* , 2 (1), 48-59.
- Stansfield, C. (1980). The cloze procedure as a progress test. *Hispania*, 63(4), 715-718.
- Steinman, L. (2002). A touch of ... class! *The Canadian Modern Language Review*, 59(2), 291.
- Stubbs, J. B., & Tucker, G. R. (1974). The cloze test as a measure of English proficiency. *The Modern Language Journal*, 58(5), 239-241.
- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism Quarterly*, Vol. 30, 1953, 415-433.
- Tenhaken, U., & Scheibner-Herzig, G. (1988). The cloze test studied from an oral communicative point of view. *Journal of Experimental Education*, 3(56), 160.
- Torres, M., & Roig, M. (2005). The cloze procedure as a test of plagiarism: The influence of text readability. *The Journal of Psychology*, 139(3), 221-31.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Wipf, J. A. (1981). Selective deletions with multiple cues: A variation of the cloze procedure as a technique for teaching and testing reading. *Die Unterrichtspraxis*, 14(2), 208-211.
- Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, 20(3), 267-293.

APPENDIX A: UN-MUTILATED TEXTS

HOW TO BE A SUCCESSFUL BUSINESSPERSON

Have you ever wondered why some people are successful in business and others are not? Here is a story about one successful businessperson. He started out washing dishes and today he owns 168 restaurants.

Zubair Kazi was born in Bhatkal, a small town in southwest India. His dream was to be an airplane pilot, and when he was 16 years old, he learned to fly a small plane.

At the age of 23 and with just a little money in his pocket, Mr. Kazi moved to the United States. He hoped to get a job in the airplane industry in California. Instead, he ended up working for a company that rented cars.

While Mr. Kazi was working at the car rental company, he frequently ate at a nearby restaurant. To save money on food he decided to get a job there. For two months, he worked as a cook's assistant. His job was to clean the kitchen and help the cook. "I did not like it," Mr. Kazi says, "but I always did the best I could."

One day, Mr. Kazi's two co-workers didn't come to work. That day, Mr. Kazi did the work of all three people in the kitchen. This really impressed the owners of the restaurant. A few months later, owners needed a manager for a new restaurant. They gave the job to Mr. Kazi. He worked hard as the manager and soon the restaurant was making a profit.

A few years later, Mr. Kazi heard about a restaurant that was losing money. The restaurant was dirty inside and the food was terrible: greasy and undercooked. Mr. Kazi borrowed money from a bank and bought the restaurant. For the first six months, Mr. Kazi worked in the restaurant from 8 a.m. to 10 p.m., seven days a week. He and his wife cleaned up the restaurant, remodeled the front of the building, and improved the cooking. They also tried hard to please the customers. If anyone had to wait more than ten minutes for their food, Mrs. Kazi gave them a free soda. Before long, the restaurant was making a profit.

A year later, Mr. Kazi sold his restaurant for a profit. With the money he earned, he bought three more restaurants that were losing money. Once again, he cleaned them up, improved the food, and re-trained the employees. Before long, these restaurants were making a profit, too.

Today Mr. Kazi owns 168 restaurants, but he is not planning to stop there. He is looking for more poorly managed restaurants to buy. "I love it when I go to buy a restaurant and find it is a mess," Mr. Kazi says. "The only way it can go is up."

Adapted From The Wall Street Journal

THE INTERVIEW

Jack comes from Denver to The Overlook Hotel to have an interview with manager Mr. Ullman. He wants to get a job at the hotel as the caretaker during the winter season when the roads to the hotel are closed due to heavy snow.

Jack: Hi! I have got an appointment with Mr. Ullman. My name is Jack Torrance.

Reception: His office is the first door on the left.

Jack: Thank you. (knocks and opens the door) Mr. Ullman? I'm Jack Torrance.

Ullman: Oh! Come in, Jack. Nice to meet you.

Jack: Nice to meet you.

Ullman: This is my secretary, Susie.

Susie: Hello!

Jack: How do you do?

Ullman: Any trouble finding here?

Jack: No trouble at all. The trip took three and a half hours.

Ullman: That is very good. Sit down a minute, Jack. Just make yourself at home. Would you like some coffee?

Jack: Well if you are going to have some, sure. Thanks.

Ullman: Susie.

...

(Susie exits, Bill enters)

...

Ullman: Get a chair Bill; sit down. Jack is going to take care of the hotel this winter. I want you to show him around the place as soon as we are done.

Bill: Fine.

Ullman: Jack is a schoolteacher.

Jack: Uh! Formerly a schoolteacher.

Bill: What do you do now?

Jack: I'm a writer. Teaching was more or less a way of making money.

Bill: This job will be quite a change for you.

Jack: Well, I'm looking for a change.

Ullman: Our people in Denver recommended Jack very highly and I agree with them. Let's see... Where were we? I was about to explain that our season runs from May 15th to October 30th. Then we close down until the following May.

Jack: Why do you do that? It seems to me that the skiing here would be fantastic.

Ullman: Oh it sure would be. The problem is the enormous cost to keep the road open. It's a 25 mile road. It gets an average of 20 feet of snow during the winter and there is no way to make it economically possible to keep it clear. When the place was built in 1907, there was very little interest in winter sports, and this place was chosen for its seclusion and beauty.

Jack: Well it's certainly got plenty of that.

Ullman: That's right. Did they give you any idea in Denver about the job?

Jack: Only in a very general way.

Ullman: Well, the winters can be very harsh. The basic idea is to cope with the damage. This consists of running the boiler, heating different parts of the hotel, and repairing damage as it occurs.

Jack: That sounds fine to me.

Ullman: Physically, it's not a very demanding job. The only thing that can get hard during the winter is a huge sense of isolation.

Jack: That's exactly what I'm looking for. I'm planning a new writing project. Five months of peace is just what I want.

Ullman: That's very good, Jack, because, for some people, solitude and isolation can become a problem.

Jack: Not for me.

Ullman: How about your wife and son? Do you think they'll like it?

Jack: They'll love it.

Ullman: Great! There is one more thing I think we should talk about. I don't want to sound melodramatic but it's something that gave a few people some doubts about the job.

Jack: I'm curious.

Ullman: I don't think they told you anything in Denver about the tragedy we had here in 1970.

Jack: I don't believe they did.

Ullman: The previous manager hired a man named Charles Grady as the winter caretaker. He came with his wife and two little girls. He had a good employment record, good references, and from what I've been told, he seemed like a completely normal person. But at some point during the winter he must have suffered a complete mental breakdown. He went crazy and killed his family with an axe. The police thought that it was a kind of claustrophobic reaction which can occur when people are shut in together over long periods of time.

Jack: That is quite a story.

Ullman: Yeah, it is. It's still hard to believe that it actually happened here but it did. I think you can understand why I wanted to tell you about it.

Jack: I certainly can. I also understand why your people in Denver left it for you to tell me.

Ullman: Obviously, some people don't like the idea of staying alone in a place where something like that happened.

Jack: You can rest assured Mr. Ullman that's not going to happen with me. And as far as my wife is concerned I'm sure she'll be absolutely fascinated when I tell her about it. She's a confirmed ghost story and horror film addict.

Adapted from the Movie: The Shining by Stanley Kubrick

APPENDIX B: THE SIX CLOZE TESTS

Group F-1 (Function words)

Aşağıdaki boşlukları **anlamli** ve **kurallı** cümleler oluşturacak şekilde doldurunuz. Boşlukların tamamı **sadece tek bir (1) kelime** ile doldurulmalıdır. Boş veya birden fazla kelime içeren cevaplar yanlış kabul edilecektir. Boşlukları anlamli ve kurallı bir biçimde doldurabilecek birden fazla farklı cevap olabilir. Bu tür bir durumda size göre en uygun cevabı seçiniz.

HOW TO BE A SUCCESSFUL BUSINESSPERSON

Have you ever wondered why some people are successful in business and others are not? Here is a story about one successful businessperson. He started out washing dishes and today he owns 168 restaurants.

Zubair Kazi was born _____ Bhatkal, a small town in southwest India. His dream was to be _____ airplane pilot, and when he was 16 years old, he learned _____ fly a small plane.

At the age of 23 and _____ just a little money in his pocket, Mr. Kazi moved to _____ United States. He hoped to get a job in the airplane industry in California. Instead, _____ ended up working for a company that rented cars.

While Mr. Kazi _____ working at the car rental company, he frequently ate _____ a nearby restaurant. To save money on food he decided to get a job _____. For two months, he worked as a cook's assistant. His job was to clean the kitchen _____ help the cook. "I did not like it," Mr. Kazi says, "_____ I always did the best I could."

One day, Mr. Kazi's two co-workers didn't come _____ work. That day, Mr. Kazi did the work of all three people in the kitchen. _____ really impressed the owners of the restaurant. A few months later, the owners needed a manager _____ a new restaurant. They gave the job to Mr. Kazi. He worked hard as _____ manager and soon the restaurant was making a profit.

_____ few years later, Mr. Kazi heard about a restaurant that _____ losing money. The restaurant was dirty inside and the food was terrible: greasy _____ undercooked. Mr. Kazi borrowed money from a bank and bought the restaurant. _____ the first six months, Mr. Kazi worked in the restaurant _____ 8 am. to 10 pm., seven days a week. He and _____ wife cleaned up the restaurant, remodeled the front _____ the building, and improved the cooking. They also tried hard _____ please the customers. If anyone had to wait more than ten minutes for _____ food, Mrs. Kazi gave them a free soda. Before long, _____ restaurant was making a profit.

A year later, Mr. Kazi sold his restaurant for _____ profit. With the money he earned, he bought three more restaurants _____ were losing money. Once again, he cleaned them up, improved the food, _____ re-trained the employees. Before long, these restaurants were making _____ profit, too.

Today Mr. Kazi owns 168 restaurants, but he _____ not planning to stop there. He is looking for more poorly managed restaurants to buy. "I love it when I go to buy a restaurant and find it is a mess," Mr. Kazi says. "The only way it can go is up."

Adapted from The Wall Street Journal

Group C-1 (Content words)

Aşağıdaki boşlukları **anlamli** ve **kurallı** cümleler oluşturacak şekilde doldurunuz. Boşlukların tamamı **sadece tek bir (1) kelime** ile doldurulmalıdır. Boş veya birden fazla kelime içeren cevaplar yanlış kabul edilecektir. Boşlukları anlamli ve kurallı bir biçimde doldurabilecek birden fazla farklı cevap olabilir. Bu tür bir durumda size göre en uygun cevabı seçiniz.

HOW TO BE A SUCCESSFUL BUSINESSPERSON

Have you ever wondered why some people are successful in business and others are not? Here is a story about one successful businessperson. He started out _____ dishes and today he owns 168 restaurants.

Zubair Kazi was born in Bhatkal, a small _____ in southwest India. His dream was to be an airplane _____, and when he was 16 years old, he learned to _____ a small plane.

At the age of 23 and with just a little _____ in his pocket, Mr. Kazi moved to the United States. He hoped to _____ a job in the airplane industry in California. Instead, he _____ up working for a company that rented cars.

While Mr. Kazi was _____ at the car rental company, he frequently ate at a nearby restaurant. To _____ money on food he decided to get a job there. For two months, he _____ as a cook's assistant. His job was to clean the kitchen and _____ the cook. "I did not like it," Mr. Kazi says, "but I always _____ the best I could."

One day, Mr. Kazi's two co-workers didn't _____ to work. That day, Mr. Kazi did the work of all three people in the _____. This really impressed the owners of the restaurant. A few months later, the owners _____ a manager for a new restaurant. They gave the job to Mr. Kazi. He _____ hard as the manager and soon the restaurant was making a _____.

A few years later, Mr. Kazi heard about a restaurant that was _____ money. The restaurant was dirty inside and the _____ was terrible: greasy and undercooked. Mr. Kazi borrowed money from a bank and _____ the restaurant. For the first six months, Mr. Kazi worked in the restaurant from 8 _____ to 10 pm., seven days a week. He and his wife cleaned up the _____, remodeled the front of the building, and improved the cooking. They also _____ hard to please the customers. If anyone had to _____ more than ten minutes for their food, Mrs. Kazi gave them a _____ soda. Before long, the restaurant was making a profit.

A year _____, Mr. Kazi sold his restaurant for a profit. With the money he _____, he bought three more restaurants that were losing money. Once _____, he cleaned them up, improved the food, and re-trained the _____. Before long, these restaurants were making a profit, too.

Today Mr. Kazi _____ 168 restaurants, but he is not planning to stop there. He is looking for more poorly managed restaurants to buy. "I love it when I go to buy a restaurant and find it is a mess," Mr. Kazi says. "The only way it can go is up."

Adapted from The Wall Street Journal

Group R-1 (Random deletion)

Aşağıdaki boşlukları **anlamli** ve **kurallı** cümleler oluşturacak şekilde doldurunuz. Boşlukların tamamı **sadece tek bir (1) kelime** ile doldurulmalıdır. Boş veya birden fazla kelime içeren cevaplar yanlış kabul edilecektir. Boşlukları anlamli ve kurallı bir biçimde doldurabilecek birden fazla farklı cevap olabilir. Bu tür bir durumda size göre en uygun cevabı seçiniz.

HOW TO BE A SUCCESSFUL BUSINESSPERSON

Have you ever wondered why some people are successful in business and others are not? Here is a story about one successful businessperson. He started out washing _____ and today he owns 168 restaurants.

Zubair Kazi was born in Bhatkal, _____ small town in southwest India. His dream was to be an airplane _____, and when he was 16 years old, he learned to fly a small _____.

At the age of 23 and with just a little money _____ his pocket, Mr. Kazi moved to the United States. He hoped to _____ a job in the airplane industry in California. Instead, he ended up _____ for a company that rented cars.

While Mr. Kazi was working at _____ car rental company, he frequently ate at a nearby restaurant. To save _____ on food he decided to get a job there. For two months, _____ worked as a cook's assistant. His job was to clean the kitchen _____ help the cook. "I did not like it," Mr. Kazi says, "but _____ always did the best I could."

One day, Mr. Kazi's two co-workers _____ come to work. That day, Mr. Kazi did the work of _____ three people in the kitchen. This really impressed the owners of the _____. A few months later, the owners needed a manager for a new _____. They gave the job to Mr. Kazi. He worked hard as the _____ and soon the restaurant was making a profit.

A few years _____, Mr. Kazi heard about a restaurant that was losing money. The restaurant _____ dirty inside and the food was terrible: greasy and undercooked. Mr. Kazi _____ money from a bank and bought the restaurant. For the first six months, Mr. Kazi _____ in the restaurant from 8 a.m. to 10 p.m., seven _____ a week. He and his wife cleaned up the restaurant, remodeled the _____ of the building, and improved the cooking. They also tried hard to _____ the customers. If anyone had to wait more than ten minutes for _____ food, Mrs. Kazi gave them a free soda. Before long, the restaurant _____ making a profit.

A year later, Mr. Kazi sold his restaurant for _____ profit. With the money he earned, he bought three more restaurants that _____ losing money. Once again, he cleaned them up, improved the food, _____ re-trained the employees. Before long, these restaurants were making a profit, _____.

Today Mr. Kazi owns 168 restaurants, but he is not planning to stop there. He is looking for more poorly managed restaurants to buy. "I love it when I go to buy a restaurant and find it is a mess," Mr. Kazi says. "The only way it can go is up."

Adapted from The Wall Street Journal

Group F-2 (Function words)

The Interview

Jack comes from Denver to The Overlook Hotel to have an interview with manager Mr. Ullman. He wants to get a job at the hotel as the caretaker during the winter season when the roads to the hotel are closed due to heavy snow.

Jack: Hi! I have got an appointment _____ Mr. Ullman. My name is Jack Torrance.

Reception: His office is the first door on the left.

Jack: Thank you. (knocks and opens the door) Mr. Ullman? I'm Jack Torrance.

Ullman: Oh! Come _____, Jack. Nice to meet you.

Jack: Nice to meet you.

Ullman: This is my secretary, Susie.

Susie: Hello!

Jack: How do you do?

Ullman: Any trouble finding here?

Jack: No trouble at all. _____ trip took three and a half hours.

Ullman: That is very good. Sit down a minute, Jack. Just make yourself _____ home. Would you like some coffee?

Jack: Well if you are going to have some, sure. Thanks.

Ullman: Susie.

...

(Susie exits, Bill enters)

...

Ullman: Get a chair Bill; sit down. Jack is _____ to take care of the hotel this winter. I want you to show him around the place as soon _____ we are done.

Bill: Fine.

Ullman: Jack is a schoolteacher.

Jack: Uh! Formerly a schoolteacher.

Bill: What do you do now?

Jack: I'm a writer. Teaching was more _____ less a way of making money.

Bill: This job will be quite a change for you.

Jack: Well, I'm looking for _____ change.

Ullman: Our people in Denver recommended Jack very highly and I agree with them. Let's see... Where were we? I was about _____ explain that our season runs from May 15th to October 30th. Then we close down _____ the following May.

Jack: Why do you do that? It seems to me that the skiing here would be fantastic.

Ullman: Oh it sure would be. _____ problem is the enormous cost to keep the road open. It's a 25 mile road. It gets an average of 20 feet _____ snow during the winter and there is no way to make it economically possible to keep it clear. When the

place _____ built in 1907, there was very little interest in winter sports, and this place was chosen for its seclusion and beauty.

Jack: Well it's certainly got plenty _____ that.

Ullman: That's right. Did they give you any idea in Denver about the job?

Jack: Only in a very general way.

Ullman: Well, the winters _____ be very harsh. The basic idea is to cope with the damage. This consists of running the boiler, heating different parts of the hotel, _____ repairing damage as it occurs.

Jack: That sounds fine to me.

Ullman: Physically, it's not a very demanding job. The only thing that can get hard _____ the winter is a huge sense of isolation.

Jack: That's exactly what I'm looking for. I'm planning a new writing project. Five months of peace is just what _____ want.

Ullman: That's very good, Jack, because, for some people, solitude and isolation can become a problem.

Jack: Not for me.

Ullman: How _____ your wife and son? Do you think they'll like it?

Jack: They'll love it.

Ullman: Great! There is one more thing _____ think we should talk about. I don't want to sound melodramatic but it's something that gave a few people some doubts about _____ job.

Jack: I am curious.

Ullman: I don't think they told you anything _____ Denver about the tragedy we had here in 1970.

Jack: I don't believe they did.

Ullman: The previous manager hired a man named Charles Grady _____ the winter caretaker. He came with his wife and two little girls. He had a good employment record, good references, and _____ what I've been told, he seemed like a completely normal person. But at some point during the winter he must have suffered _____ complete mental breakdown. He went crazy and killed his family with an axe. The police thought that it was a kind of claustrophobic reaction _____ can occur when people are shut in together over long periods of time.

Jack: That is quite a story.

Ullman: Yeah, it is. It's still hard to believe _____ it actually happened here but it did. I think you can understand why I wanted to tell you about it.

Jack: I certainly can. _____ also understand why your people in Denver left it for you to tell me.

Ullman: Obviously, some people don't like the idea of staying alone in a place _____ something like that happened.

Jack: You can rest assured Mr. Ullman that's not going to happen with me. And as far as _____ wife is concerned I'm sure she'll be absolutely fascinated when I tell her about it. She's a confirmed ghost story and horror film addict.

Adapted from the Movie: The Shining by Stanley Kubrick

Group C-2 (Content words)

The Interview

Jack comes from Denver to The Overlook Hotel to have an interview with manager Mr. Ullman. He wants to get a job at the hotel as the caretaker during the winter season when the roads to the hotel are closed due to heavy snow.

Jack: Hi! I have got an _____ with Mr. Ullman. My name is Jack Torrance.

Reception: His office is the first door on the left.

Jack: Thank you. (*knocks and opens the door*) Mr. Ullman? I'm Jack Torrance.

Ullman: Oh! _____ in, Jack. Nice to meet you.

Jack: Nice to meet you.

Ullman: This is my secretary, Susie.

Susie: Hello!

Jack: How do you do?

Ullman: Any trouble _____ here?

Jack: No trouble at all. The trip took three and a half hours.

Ullman: That is very good. Sit down a minute, Jack. Just _____ yourself at home. Would you like some coffee?

Jack: Well if you are going to have some, sure. Thanks.

Ullman: Susie.

...

(Susie exits, Bill enters)

...

Ullman: _____ a chair Bill; sit down. Jack is going to take care of the hotel this winter. I want you to _____ him around the place as soon as we are done.

Bill: Fine.

Ullman: Jack is a schoolteacher.

Jack: Uh! Formerly a schoolteacher.

Bill: What do you _____ now?

Jack: I'm a writer. Teaching was more or less a way of making money.

Bill: This job will be quite a change for you.

Jack: Well, I'm _____ for a change.

Ullman: Our people in Denver recommended Jack very highly and I agree with them. Let's see... Where were we? I was about to explain that our season _____ from May 15th to October 30th. Then we close down until the following May.

Jack: Why do you do that? It _____ to me that the skiing here would be fantastic.

Ullman: Oh it sure would be. The problem is the enormous cost to _____ the road open. It's a 25 mile road. It gets an average of 20 feet of snow during the _____ and there is no way to make it economically possible to keep it clear. When the place was built in 1907, there was very little _____ in winter sports, and this place was chosen for its seclusion and beauty.

Jack: Well it's certainly _____ plenty of that.

Ullman: That's right. Did they give you any idea in Denver about the job?

Jack: Only in a very general _____.

Ullman: Well, the winters can be very harsh. The basic idea is to cope with the damage. This consists of running the boiler, heating different parts of the hotel, and _____ damage as it occurs.

Jack: That sounds fine to me.

Ullman: Physically, it's not a very demanding job. The only thing that can get hard during the winter is a huge _____ of isolation.

Jack: That's exactly what I'm looking for. I'm planning a new writing project. Five months of peace is just what I _____.

Ullman: That's very good, Jack, because, for some people, solitude and isolation can become a problem.

Jack: Not for me.

Ullman: How about your _____ and son? Do you think they'll like it?

Jack: They'll love it.

Ullman: Great! There is one more thing I think we should _____ about. I don't want to sound melodramatic but it's something that gave a few people some doubts about the _____.

Jack: I'm curious.

Ullman: I don't think they told you anything in Denver about the tragedy we had here in 1970.

Jack: I don't _____ they did.

Ullman: The previous manager hired a man named Charles Grady as the winter caretaker. He came with his wife and two little girls. He _____ a good employment record, good references, and from what I've been told, he seemed like a completely normal person. But at some _____ during the winter he must have suffered a complete mental breakdown. He went crazy and killed his family with an axe. The police thought that it was a _____ of claustrophobic reaction which can occur when people are shut in together over long _____ of time.

Jack: That is quite a story.

Ullman: Yeah, it is. It's still hard to _____ that it actually happened here but it did. I think you can understand why I _____ to tell you about it.

Jack: I certainly can. I also understand why your people in Denver left it for you to tell me.

Ullman: Obviously, some _____ don't like the idea of staying alone in a place where something like that happened.

Jack: You can rest assured Mr. Ullman that's not going to _____ with me. And as far as my wife is concerned I'm sure she'll be absolutely fascinated when I tell her about it. She's a confirmed ghost story and horror film addict.

Adapted from the Movie: The Shining by Stanley Kubrick

Group R-2 (Random deletion)

The Interview

Jack comes from Denver to The Overlook Hotel to have an interview with manager Mr. Ullman. He wants to get a job at the hotel as the caretaker during the winter season when the roads to the hotel are closed due to heavy snow.

Jack: Hi! I have got an appointment with Mr. Ullman. My name is Jack Torrance.

Reception: His _____ is the first door on the left.

Jack: Thank you. (*knocks and opens the door*) Mr. Ullman? I'm Jack Torrance.

Ullman: Oh! Come in, Jack. Nice to meet _____.

Jack: Nice to meet you.

Ullman: This is my secretary, Susie.

Susie: Hello!

Jack: How do you do?

Ullman: Any trouble finding here?

Jack: No trouble at _____. The trip took three and a half hours.

Ullman: That is very good. Sit down a minute, Jack. Just make yourself at _____. Would you like some coffee?

Jack: Well if you are going to have some, sure. Thanks.

Ullman: Susie.

...

(Susie exits, Bill enters)

...

Ullman: Get a chair Bill; sit _____. Jack is going to take care of the hotel this winter. I want you to show him around the place as _____ as we are done.

Bill: Fine.

Ullman: Jack is a schoolteacher.

Jack: Uh! Formerly a schoolteacher.

Bill: What do you do now?

Jack: I'm a writer. _____ was more or less a way of making money.

Bill: This job will be quite a change for you.

Jack: Well, I'm looking _____ a change.

Ullman: Our people in Denver recommended Jack very highly and I agree with them. Let's see... Where were we? I _____ about to explain that our season runs from May 15th to October 30th. Then we close down until the following May.

Jack: _____ do you do that? It seems to me that the skiing here would be fantastic.

Ullman: Oh it sure would be. The _____ is the enormous cost to keep the road open. It's a 25 mile road. It gets an average of 20 feet _____ snow during the winter and there is no way to make it economically possible to keep it clear. When the place _____ built in 1907, there was very little interest in winter sports, and this place was chosen for its seclusion and beauty.

Jack: Well _____ certainly got plenty of that.

Ullman: That's right. Did they give you any idea in Denver about the job?

Jack: Only in a very _____ way.

Ullman: Well, the winters can be very harsh. The basic idea is to cope with the damage. This consists of running the boiler, _____ different parts of the hotel, and repairing damage as it occurs.

Jack: That sounds fine to me.

Ullman: Physically, it's not a very _____ job. The only thing that can get hard during the winter is a huge sense of isolation.

Jack: That's exactly what I'm _____ for. I'm planning a new writing project. Five months of peace is just what I want.

Ullman: That's very good, Jack, because, _____ some people, solitude and isolation can become a problem.

Jack: Not for me.

Ullman: How about your wife and son? Do you think _____ like it?

Jack: They'll love it.

Ullman: Great! There is one more thing I think we should talk about. I don't want to _____ melodramatic but it's something that gave a few people some doubts about the job.

Jack: I'm curious.

Ullman: I don't think they told _____ anything in Denver about the tragedy we had here in 1970.

Jack: I don't believe they did.

Ullman: The previous manager hired a _____ named Charles Grady as the winter caretaker. He came with his wife and two little girls. He had a good employment _____, good references, and from what I've been told, he seemed like a completely normal person. But at some point during the _____ he must have suffered a complete mental breakdown. He went crazy and killed his family with an axe. The police thought _____ it was a kind of claustrophobic reaction which can occur when people are shut in together over long periods of time.

Jack: _____ is quite a story.

Ullman: Yeah, it is. It's still hard to believe that it actually happened here but it did. I _____ you can understand why I wanted to tell you about it.

Jack: I certainly can. I also understand why your people in _____ left it for you to tell me.

Ullman: Obviously, some people don't like the idea of staying alone in a place where _____ like that happened.

Jack: You can rest assured Mr. Ullman that's not going to happen with me. And as far as my wife is concerned I'm sure she'll be absolutely fascinated when I tell her about it. She's a confirmed ghost story and horror film addict.

Adapted from the Movie: The Shining by Stanley Kubrick

APPENDIX C: ACCEPTABLE ANSWERS

Test F-1

- 8 in
13 he, it
19 after, throughout
24 the
25 his
27 which

Test C-1

- 1 cleaning
2 village, city
3 designer
4 design, make
5 cash, bread
6 find, have
13 want, go
14 restaurant
20 redecorated, improved
22 kitchen
23 worked
25 cold, fresh, wonderful, lemon, fruit
26 ago

27 had
29 workers, crew

30 has

Test R-1

1 up

4 vehicle

6 find, have

13 hadn't

14 the, these

15 company, place

16 company

17 boss

20 got

21 stayed, was

23 walls, decoration, appearance, inside, design

24 serve, interest, satisfy

25 the

26 started, began

30 again

Test F-2

3 my
5 here, ready, expecting, about
8 some
10 through
11 our
16 also
17 in
18 we
20 we
21 this
26 that

Test C-2

1 interview
5 take, find
6 take
8 ready, waiting, searching, here
9 is, lasts
12 season, day
17 amount
18 need
19 daughter

- 20 think
- 21 hotel, isolation
- 22 think
- 24 time
- 27 accept, admit
- 28 want, have, need

Test R-2

- 1 room
- 5 please
- 7 that
- 11 reason
- 16 renewing, in
- 17 hard, difficult, bad
- 18 waiting
- 19 with
- 21 be
- 23 worker, person
- 25 season
- 27 it, this
- 28 hope, guess
- 30 things