

Jimma University
College of Social Sciences and Humanities
Department of English Language and Literature
MA in Teaching English as a Foreign Language (TEFL)

**Assessment of the Construct-irrelevant Factors that Affect Students’
Reading English Comprehension Test: The Case of Sire and
Ano Preparatory Schools of Grade 11 Students**

By
Lule Moroda

**A Thesis Submitted in Partial Fulfillment of the Requirements for
Master of Arts in TEFL**

August 2015

**Assessment of the Construct–irrelevant Factors Affecting Students’
Reading English Comprehension Test: The Case of Sire and
Ano Preparatory Schools of Grade 11 Students**

By

Lule Moroda

**A Thesis Submitted in Partial Fulfillment of the Requirements for
Master of Arts in TEFL**

Department of English Language and Literature

College of Social Sciences and Humanities

Jimma University

August 2015

Declaration, Confirmation, Approval and Evaluation

Research Title: Assessment of the Construct-irrelevant Factors that Affect Students' Reading English Comprehension Test

Declaration

I, the undersigned, declare that this thesis is my original work, not presented for any degree in any universities, and that all the sources used for it are duly acknowledged.

_____	_____	_____
Name	Signature	Date

Confirmation and Approval

This thesis has been submitted for examination with my approval as a thesis advisor.

Principal Advisor:

_____	_____	_____
Name	Signature	Date

Co-Advisor:

_____	_____	_____
Name	Signature	Date

Thesis Evaluators

_____	_____	_____
Principal Advisor	Signature	Date

_____	_____	_____
Co-advisor	Signature	Date

_____	_____	_____
External Examiner	Signature	Date

_____	_____	_____
Internal Examiner	Signature	Date

_____	_____	_____
Chairperson	Signature	Date

Acknowledgments

First of all, I would like to thank the Almighty and Omniscient God who gave me the fate to pursue my MA in TEFL study at Jimma University where I gained much. Many persons directly or indirectly helped me during my study. I would like, therefore, to express my sincere gratitude to all these people for helping me to accomplish so much. Particularly, I would like to express my gratefulness to the following wonderful people: **Dr. Temesgen Mereba**, my major advisor, who helped me from the early design of the research proposal by providing me constructive comments and guidance with quick response within the time frame designed for this thesis work. I also thank **Dr. Tesfaye Gebeyehu**, my co-advisor deeply from my heart for his constructive comments and guidance with quick response within the time frame designed for this work. In fact, the completion of this work would not have been possible without their guidance, invaluable suggestions, comments, professional experts and constructive criticisms. They had taken great care to improve the content of this thesis.

My greatest debt of gratitude in preparation of this paper is to all my respondents for their time and interests. They gave me a small portion of their life. In addition, it is pleasant to acknowledge Ato Dessalegn Getachew for his constructive suggestions and helping me in operating the SPSS data analysis and arranging the typing format of the manuscript of this thesis.

My special thanks go to my brother Ebsa Moroda (LLB) for his continuous guidance, help, moral support from the early stages of this work were immemorial.

The most important person in this effort has been Ato Soressa Mekonnen who provided me material support and the support in commenting and arranging the statistical analysis of data concerning tests.

Abstract

The purpose of this study was to assess the construct-irrelevant factors affecting students' performance in reading English comprehension test at Ano and Sire Preparatory Schools of Grade 11. Specifically, it was designed to address the major construct-irrelevant factors, the adverse effects of those factors and its strategies to overcome it. Therefore, the Mixed Research Design method was used in conducting this study. In order to gather accessible data, 44 students were selected by systematic random sampling and they filled and answered the checklist questionnaire and test respectively. Moreover, to crosscheck the data obtained via test and questionnaire, 5 English teachers were selected by purposive sampling technique and were interviewed. In addition, the document analysis from the exam-bank that was prepared in 2014/15 by the subject teachers was analyzed qualitatively. The quantitative data obtained through tests were analyzed using SPSS version 20 to compute the descriptive statistics like mean, range, standard deviation, percentage, correlation, inter-rater reliability (Kappa) and the inferential statistics like the paired sampled T-test. On the other hand, the qualitative data obtained via students' suggestion from checklist questionnaire was analyzed qualitatively. Finally, the findings of the study indicated that students' performance on reading comprehension test found to be affected by unfamiliar languages, ineffective use of test-accommodations, and unfamiliarity of students to test items, and the badly written texts and test questions which in turn have an adverse effect of test-anxiety, feeling of getting panic and worry for low test scores, loss of interest and hatred of the subject teacher. Eventually, sufficient allocation and provision of test materials, careful test preparation and the usage of computer skilled persons were suggested as the major strategies used to overcome the construct-irrelevant factors in reading English comprehension classroom test.

Table of Contents

Abstract	iii
Table of Contents	vii
List of Tables	vii
Abbreviations Used	viii
INTRODUCTION: CHAPTER ONE	13
1.1. Background of the Study	13
1.2. Statement of the Problem	16
1.3. Objective of the Study	19
1.3.1. The Main Objective of the Study	19
1.3.2. Specific Objectives	19
1.4. Research Questions	19
1.5. Significance of the Study	20
1.6. Limitation of the Study	20
1.7. Delimitation of the Study	10
1.8. Definitions of Key Terms	22
CHAPTER TWO: REVIEW OF RELATED LITERATURE	28
2.1. What is Reading?	28
2.2. Reading and Comprehension	14
2.3. The Importance of Reading	14
2.4. What is Test?	15
2.5. Purpose of Tests	16
2.6. Construct-Irrelevant Factors	20
2.6.1. The Language Test Performance in Reading Comprehension Test	21
2.6.2. The Factors Influencing Students' Performance in English Classroom Tests	22
2.6.2.1. The Factors Related to Language	393
2.6.2.2. The Factors Related to the Ineffective Use of Test Accommodations	393
2.6.2.2.1. Two Level Taxonomy for Organizing Test Accommodations	404
2.6.2.3. Students' Unfamiliarity with the Test and/or Question Format	404
2.6.2.4. The Factors Related to Badly Typed Texts and Test-Questions	40

2.7. Strategies to Alleviate the Challenges of Construct-irrelevant Factors	415
2.7.1. Language	415
2.7.1.1. Use Simple Grammar and Sentence Structure	415
2.7.1.2. Use Common Vocabulary Wherever Possible	426
2.7.2. Test Accommodations	426
2.7.2.1. Utilize Accommodations that do not affect what is being measured	437
2.7.2.2 Use Accommodations that are Students Responsive	437
2.7.3. Students' Unfamiliarity with the Test and / or Question Format	448
2.7.3.1. Align Testing with Instructional Practice	448
2.7.3.2. Teach Students how to Negotiate Different Item/Test Formats	448
2.7.3.3. Ensure that Students Know the "Rules of the Testing Game"	459
2.7.3.4. Allow Students to Experience the Testing Conditions	459
2.7.3.5. Teach Specific Test Taking Techniques	459
2.7.4. The Badly Typed and Usage of Too Much Text in Too a Small Space	30
CHAPTER THREE: RESEARCH METHODOLOGY	31
3.1. Design of the Study	31
3.2. Participants of the Study	32
3.3. Sample Size and Sampling Techniques	33
3.4. Data Collection Instruments	34
3.4.1. Test	35
3.4.2. Questionnaire	848
3.4.3. Semi- structured Interview	869
3.4.4. Document Analysis	40
3.5. Data Collection Procedures	40
3.6. Data Analysis Procedures	41
3.7. Ethical Considerations	41
CHAPTER FOUR: FINDINGS AND DISCUSSION	43
4.1. Construct-irrelevant Factors Influencing Students' Reading Comprehension Test	44
4.2. The Organization of Data Obtained from the Four Major Instruments	44
4.3. The Factors Related to Language	47
4.3.1. The Factors Related to Questions Containing Unfamiliar Languages	54
4.4. The Factors Related to Test Accommodations Influencing Students' Performance	55
4.5. The Item/test Format Factors on Students' Language Test Performance	58

4.6. The Factors Related to Badly-typed Texts and Test-questions	59
4.7. The Adverse Effects of Unfamiliar Languages on Students' Test Performance	63
4.8. The Adverse Effects of Test-accommodation in Reading Comprehension Test	64
4.9. The Adverse Effects of Unfamiliar Item/test Format	65
4.10. The Adverse Effects of the Badly-typed Texts/Test- Questions	67
4.11. The Strategies to overcome the Language Factors in Reading Comprehension Test	68
4.12. The Strategies to overcome the Problems of Test-accommodations Factors	70
4.13. Strategies for coping with the Problems Related to Item/Test Format	71
4.14. Strategies for coping with the Problems of Badly Typed Texts and Test Questions	72
CHAPTER FIVE: SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS	74
5.1. Summary	74
5.2. Conclusions	76
5.3. Recommendations	77
REFERENCES	79
APPENDIX A: Questionnaire to be Filled by Students	85
APPENDIX B: Gaaffilee Chekliistiin Barattotaaf Qophaa'e	90
APPENDIX C: Reading Comprehension Test for Students	95
APPENDIX D: Reading Comprehension Test for Students	98
APPENDIX E: Semi-structured Interview for Teachers	101
APPENDIX F: Document Analysis from the Exam Bank	103
APPENDIX G: Percentages of Student Responses with Different Construct-irrelevant Variables	104
APPENDIX H: Student Test Scores in Reading Comprehension Test	107
APPENDIX I: The Interpretation of Proportions with Equivalent Percentages	109
APPENDIX J: The Sample Transcription of Student Suggestions from Checklist Questionnaire	110
APPENDIX K: Sample Transcription of Teacher's Interview	114
APPENDIX L: Sample Transcription of Exam-paper Analysis	117
APPENDIX M: Model Test Sample from Ano Preparatory School	119

List of Tables

Table 4.2.1. Questions Analyzed under Different Construct-irrelevant Factors	925
Table 4.2.2. Questions Analyzed under the Adverse Effects of Construct-irrelevant Factors	925
Table 4.2.3. Questions Analyzed as Strategies to Overcome Construct-irrelevant Factors	936
Table 4.3.1. Students' Test Scores Rated by the Scorer "A"	947
Table 4.3.2. Students' Test Scores Rated by Scorer "B"	958
Table 4.3.3. Descriptive Statistics for the Test Results of Students	969
Table 4.3.4. The Symmetric Measures of Inter-rater Reliability of the Two Scorers	51
Table 4.3.5. Correlation of the Test Scores of Unfamiliar Languages and Test Languages	52
Table 4.3.6. The Results of Paired Sampled T- test between Unfamiliar and Familiar Language Tests	53
Table 4.3.7.1. Difficult Language Factors in Reading Comprehension Test Questions	54
Table 4.4.1. The Test-accommodation Related Factors in English Language Test	56
Table 4.5.1. The Factors Related to Students' Unfamiliarity with Item/test Format	58
Table 4.6.1. The Badly Typed Texts and Test Questions Factors	60
Table 4.7.1. The Adverse Effects of Unfamiliar Languages in Reading Comprehension	64
Table 4.8.1. The Adverse Effects of Test-accommodation	65
Table 4.9.1. Students' Lack of Experience in Answering Short-answer Questions	66
Table 4.10.1. The Adverse Effects of the Badly Typed Texts/Test Questions	67
Table 4.11.1. Strategies used by Test-developers to overcome the Language Problems	68
Table 4.12.1. The Strategies used by Test-developers overcome the Inappropriate usage of Test-accommodations	70
Table 4.13.1. The Strategies to overcome Item/Test Format Factors by Test-developers	71
Table 4.14.1. Strategies used by Test-developers to overcome the Badly Written Texts and Test-questions	72

Abbreviations Used

CLA: Communicative Language Ability

ELL: English Language Learners

ICDR: Institute for Curriculum Development and Research

MOE: Ministry of Education

JUCDE: Jimma University Continuing and Distance Education

SPSS: Statistical Package for Social Sciences

CHAPTER ONE

1. INTRODUCTION

There are many factors that affect language test performance in English language test. Among the many factors, the construct-irrelevant factors which is the factors that the students encounter on tests beyond the skill, knowledge, and ability that the test is intended to measure is one. Therefore, this chapter presents the background of the study, statement of the problem, objective of the study, research questions, limitation of the study, delimitation of the study, significance of the study, and definitions of key terms concerning the construct-irrelevant factors to students' performance in reading English comprehension test.

1.1. Background of the Study

Factors affecting language test performance are varied and complex. In a unified model, Bachman (1990, as cited in Bachman & Palmer, 1996) sought to explain language test performance and set out four major influences on language test scores: communicative language ability (CLA), test method facets, individual characteristics, and random measurement error or unreliability. With reference to his unified model, Bachman (2000) grouped research involving factors affecting language test performance into three intertwined areas: the characteristics of the testing procedures (as relevant to test method facets), characteristics of test-takers themselves (as relevant to CLA and test-taker characteristics), and the process and strategies used by test-takers in response to test-tasks (as an interaction among CLA, test-method facets, and test-taker characteristics).

The communicative language ability that affects students test performance includes language competence, strategic competence, and sociolinguistic competence (Hughes, 1989). Test method facets such as environment, test rubric, input, expected response, and the relationship between input and response affects students' performance on language test. In addition, the attributes of the test taker that are not considered part of the language abilities the teacher wants to measure such as cultural background, knowledge, cognitive

abilities, sex, age are also factors that affect language test performance. And finally, random factors such as the interaction among the components of communicative language abilities, test method facets, personal characteristics and measurement error influences students test performance (Bachman, 1990).

All these are the factors that affect students' test performance in relation to both construct-relevant and construct-irrelevant factors for all language learners including ELLs and native language learners. According to Bachman (1990), a student's performance on any language test is influenced by a large number of factors that must be taken into account in the construction and development of language tests. Accordingly, Bachman(1990) and later Bachman and Palmer(1996) presented a theory of language testing that contained not only different aspects of language ability but also the methods and other factors involved in the measurement of this ability. Individual's performance on language test may vary due to the influence of one's own language ability, test methods facets and other factors. In other words, student's performance on language tests may vary due to the influence of construct- relevant and construct-irrelevant factors.

According to Hughes (1989), construct refers to any underlying ability (or trait) which is hypothesized in a theory of language ability. The construct-irrelevant factors, on the other hand, are the factors that affect language test performance beyond the ability, knowledge, and skills that the test is intended to measure, yet the most important, researched aspects of tests. But in recent years, great interest has arisen in the study of the challenges to students' test performance on greater test scores in the structured and unstructured test with construct-irrelevant factors which is the issues beyond the language knowledge, skills and abilities that the test is intended to measure (Abedi, 2002).

Language knowledge which is an information specific to language use that is stored in memory includes knowledge of syntax, knowledge of rhetorical organization, knowledge of cohesion, knowledge of register; language skills consist of listening, reading, speaking, and writing; and language ability consist of the four skills and several components such as grammar, vocabulary, and pronunciation (Bachman & Palmer, 1996). But, language

skill is not being part of language ability at all, but to be the contextualized realization of the ability to use language in the performance of specific language use tasks (Bachman & Palmer, 1996).

Apart from the construct-relevant factors that affect students' test performance, the extraneous variables such as reading difficulty, difficult vocabulary, vagueness, and ambiguity, long and complex sentences are grouped under the construct-irrelevant factors affecting reading comprehension test.

The most important consideration in designing and developing a language test is the use for which it is intended (Bachman & Palmer, 1996). That is the ability that the test is intended to measure. Bachman(1990, p.81) states; "If we are to develop language tests appropriately, for the purposes for which they are intended, we must base them on clear definitions of both abilities we wish to measure and the means by which we observe and measure these abilities."

Students' test performances in classroom tests are low due to a number of factors beyond the knowledge, skills, and abilities that the test is intended to measure (Fairbairn, 2007). Matters of language, ineffective use of test accommodations, unfamiliarity of students with the test or question format and badly written and/or too much text in too small a space may hinder student performance by introducing and/or falling short of eliminating construct-irrelevant variant factors. That is, these factors may impede students' performance in such a way that the result of the performance represent issues beyond the content/or skills that are the focus of the test.

It is therefore, this study was intended to assess the factors that affect students' performance in reading comprehension test in relation to the construct-irrelevant variant factors and to suggest the recommendations for the problems encountered in Ano, and Sire preparatory schools of Grade 11.

1.2. Statement of the Problem

In an age of test-based accountability, accurate assessment is paramount. When testing language learners (L₂ students), challenges associated with language in test questions, the inappropriate usage of test accommodation, the unfamiliarity of students to the test/item format, and badly typed texts can undermine this accuracy (Fairbairn, 2007).

In relation to the above challenges, among the many principles that are considered in test construction, one is that a test should be free of extraneous factors that affect students test performance. That is, reading difficulty, ambiguity, vague, long and complex sentences, difficult vocabulary, etc. (Thorndike et al., 1977; Ebel, 1979; Gronland, 1981).

Low students test performance result in serious consequences for students and teachers alike, particularly in an era strongly focused on test based accountability (Fairbairn, 2007). These negative results include misunderstandings about the ability, knowledge, skills possessed by students and the carelessness on the construction and development of tests in relation to the extraneous factors; claims that teachers and schools are ineffective; and feelings of frustration on the part of educators, their students, and family and community members. Thus, a number of problems keep classroom tests from being accurate measures of students' performance such as ambiguous and unclear test-questions; and the tests are too short to provide an adequate sample of the body of content to be covered that introduce undue error and are not fair to the test-takers.

Therefore, language issues are the main concern when testing students. Menken (2005) realizes that the language used on tests is different from everyday language that most L₂ students quickly learn and even from the academic language used in instructional settings.

Test accommodations do offer some promise for helping to level the playing field for L₂ students but many commonly used accommodations are fraught with their own challenges (Abedi, 2002; Bailey, 2007). Some accommodations, such as extra-time, are

beneficial to all students in non-standardized test but when they are used excessively, it cannot provide a positive effect on tests of standardized tests which is highly time dependent. That is, without extra-time.

Students' unfamiliarity with the test and/or question format is a third issue deserving of attention when testing L₂ students. Students who are unfamiliar with the format of the test itself or test question may likely have difficulty in demonstrating what they know and are able to do (Hughes, 1989). Students are familiar with oral questions and multiple-choice questions in reading comprehension test but they are not experienced open-ended questions on tests.

The badly typed and the usage of too much text in too a small space is the fourth issue which affects students test performance by causing anxiety on the side of students (Hughes,1989). Mostly reading passages are more vulnerable to this challenge more than the other skills.

Individual's performance on language test may vary due to the influence of the language ability, test method facets and other factors (Hacuta, 2009). Highly relying on the student's ability, skills and knowledge by leaving behind the construct-irrelevant factors causes students' poor performance (Bachman & Palmer, 1996). Thus, the inability to use both construct-relevant and irrelevant factors simultaneously by the test developer causes students poor language test performance.

Teachers complaining about reading comprehension text and test question correction orientation while the students are doing the test in the classroom, and the sanctions imposed on the usage of test materials by the school administrative bodies in the allocation of budget for the test materials are the main problems that initiated the researcher to conduct the research on this issue. These are the reasons for the researcher to see at the construct-irrelevant factors influencing students' performance on language test particularly in reading comprehension test at Ano and Sire preparatory schools of Grade 11.

Many researches that were conducted on construct-irrelevant variation factors to students' (ELLs) performance on language test are reviewed and the researcher found that many of them were conducted in western countries particularly in the USA. This was mainly focused on the general language test. But the local research conducted in Ethiopian cases mainly focuses on ability, skill and knowledge that the test is intended to measure on reading comprehension test. For example, Biruk Abera (2013) conducted his research on assessing teacher and student related factors that help improve students reading comprehension for academic purpose in the case of Yehibret Fire high school of Grade ten students. Thus, he found that factors affecting reading English comprehension test that are related to students includes weak decoding skills, poor reading prosody, and weak language processing skills, difficult vocabulary in context, memory and lack of active processing. The teacher related factors such as teaching strategies and the type of text(descriptive, expository, argumentative and narrative) which are not addressed the construct-irrelevant factors that affect students performance on reading comprehension test beyond the ability, skill, and knowledge that the test is intended to measure.

In addition, Tarekegn Wondimu (2012) conducted research on a comparative study of reading comprehension skill between governmental and non-governmental preparatory school students of Grade 12 in Jimma using test, questionnaire, and interview and found that students' background knowledge or schemata, the type of text (story), interest and teaching strategies were found to have an influence on students' performance. This is again cannot address the challenges of students' performance on reading comprehension test beyond the ability, skill, and knowledge that the test is intended to measure. However, this study focused on the problems to students' performance on reading comprehension test in the sample study beyond the ability, skill and knowledge that the test is intended to measure.

The researcher, therefore, liked to check if the construct-irrelevant factors to students' performance on reading comprehension test affect their language test performance and assess the extent to which those factors influence the target language test in English in the Ano and Sire preparatory schools context, so that this study could fill the gap.

1.3. Objective of the Study

1.3.1. The Main Objective of the Study

The objective of this study was to assess the major construct-irrelevant factors that affect students' reading English comprehension test with the particular reference to Grade 11 students at Ano and Sire Preparatory Schools in Western Ethiopia, Eastern Wollega Zone, Oromia Regional State.

1.3.2. Specific Objectives

The specific objectives of this study are:

1. To identify the major construct-irrelevant factors affecting students' performance in reading English comprehension test.
2. To assess the adverse effects of the construct-irrelevant factors influencing students' performance in reading English comprehension test.
3. To find out strategies used by test-developers to overcome the construct-irrelevant factors to students' test performance in reading English comprehension test.

1.4. Research Questions

The following are research questions that were answered in the study:

1. What are the major construct-irrelevant factors affecting students' performance in reading English comprehension test?
2. What are the adverse effects of the construct-irrelevant factors that influence students' performance while they are taking reading English comprehension test?
3. What are the strategies used by teachers or test developers to overcome the construct-irrelevant factors to students' performance in reading English comprehension test?

1.5. Significance of the Study

The result of this study may be helpful for teachers or test developers who have been complaining why their students' test performance is low in reading English comprehension test. As this study aimed to indicate the construct-irrelevant factors to students' performance on the target test and the strategies to overcome it, they may come to know the source of construct-irrelevant factors of students' performance on test from this study. This leads teachers or test developers to create students' high performance on tests in line with the goal of their testing programs and actual test practices in the classroom tests by identifying their students level of test performance using the sample reading comprehension test.

It may also be indirectly useful for curriculum designers, textbook writers and large-scale test developers in developing the test free of extraneous factors and identify a means of lowering the construct-irrelevant challenges of a test if the finding of this study is available from the internet. Moreover, students who are taking the open-ended reading comprehension test would be benefited by the solution that were possibly be suggested by the study. It was revealed the effects of construct-irrelevant variation factors on someone's test result achievement.

By this, we may know that it could give awareness both for students and teachers and encourage them to work on how to overcome such factors on reading comprehension test without changing the construct being measured. That is the ability, skill and knowledge that the test is intended to measure. Finally, this study may benefit researchers who want to conduct further investigation on topics related to this research.

1.6. Limitations of the Study

This study was believed to have certain limitations. Primarily, it is worth keeping in mind that the sample size of the study was limited to only 5 teachers and 44 students from the two schools. This could often affect the generalizability of the study. It was better and

more effective if a good number of students and teachers were included in the study to gather sufficient data to obtain better results.

The researcher also recognizes that demonstrating all the construct-irrelevant factors was not carried out within a short period of time and required an intensive investigation to reflect all the problems. Therefore, the instruments that were used to gather data was limited to raise only some of the construct-irrelevant factors on reading comprehension test.

The other problem for the researcher is the failure of the sample respondents to answer the open-ended reading comprehension test such as fatigue, a bad day and temporary illness which are another factor that cause results might not accurately reflects the performances of all the members of the included population. In the consideration of such conditions, the model test assumed to be presented in the situation when the climate of the school compound is relatively conducive in the morning.

Moreover, when the same version of a test with different languages (difficult vocabularies and expressions) was administered twice to the same respondents, it might expose respondents to infer the answer from the first session through practice effect and memory in which the test-takers retake the test, there was the possibility of getting some answers correct. Therefore, it was better to elongate the time interval between the two sessions.

Finally, analyzing the open-ended questions leads to the subjective interpretation of students' suggestions and responses by the researcher from the checklist questionnaire. Therefore, it was advisable if the questions were presented with the closed-ended questions.

1.7. Delimitations of the Study

Factors affecting students' performance in reading English comprehension test are varied and complex. Therefore, it is very difficult to include all the factors and carry out the study within short period of time. Thus, this study was aimed to assess the challenges to students' performance in relation to construct-irrelevant factors that cause test-takers poor performance in reading comprehension test beyond the ability, skill and knowledge that the test is intended to measure at Ano and Sire Preparatory Schools of Grade 11. Therefore, the sample size of the study was delimited to 5 teachers and 44 students and two schools namely Ano and Sire Preparatory Schools.

Each units of Grade 11 students' English textbook provides reading passages followed by open-ended items and sometimes closed-ended reading comprehension questions to be answered orally by the students. In addition, the classroom based tests provides the test-takers with reading comprehension tests using multiple-choice items that invite them to guess the answer. Therefore, this study was then aimed to assess the construct-irrelevant factors that affect students' performance in short-answer item of reading comprehension test at this grade level.

1.8. Definitions of Key Terms

Accommodations: changes to testing procedures including presentation of test materials, students' response to test-items, invigilators test condition facilitation, scheduling and test setting and structure of the test.

Accountability: being responsible for one's own decisions or actions in designing, constructing, administering, and scoring the test.

Assessing: making a judgment about the nature of construct-irrelevant factors challenging students' performance on reading comprehension test.

Comprehension: the understanding and interpretation of what one reads.

Construct: the ability that the test is intended to measure based on the theoretical framework of a test.

Construct-irrelevant factors: the factors that influence students' language test performance beyond the ability, skill and knowledge that the test is intended to measure.

Construct-relevant factors: the factors that influence students' language test performance based on the ability, skill and knowledge that the test is intended to measure.

Direct linguistic support accommodations: involve adjustments to the language of the test.

Indirect linguistic support accommodations: involve adjustments to the conditions under which a test is administered.

Performance: the act that involves effort to do the test.

Testing accommodations: involve the changes in assessment environment that do not fundamentally alter the construct being assessed.

CHAPTER TWO

2. REVIEW OF RELATED LITERATURE

2.1. What is Reading?

In foreign language learning, reading is likewise a skill that teachers simply expect learners to acquire. Most formal tests use the written word as a stimulus for test-taker response. Reading might be tested within a content focused battery: texts carry meaning for readers, that interest them, that related to their academic background, leisure interests, intellectual levels and so on, might motivate a deeper reading than the traditional, relatively inoffensive or even countless texts (Alderson, 2000). Reading, arguably the most essential skill for success in all educational contexts, remains a skill of paramount importance as we create assessments of general language ability (Brown, 2004).

Reading comprehension consists of the extraction and construction of meaning from the interaction with a written text (Hooves & Gough, 1990). This definition entails ideas that readers play an active role in comprehending because they gather the meaning that the text explicitly conveys and construct their own meaning on their background and at least two entities are involved in comprehension: the reader and the text. The simple view of reading (Hoover & Gough, 1990) proposes that reading comprehension is the product of accurate identification of the printed words (decoding) and the semantic and syntactic relationships among the words and phrases (linguistic comprehension).

Readers are expected to interact with the text to different degrees. Each task involved in comprehending a written text requires a different level of cognitive processing and demands the use of different sources of information (Basbaraba, Yonanoff, Alozo, and Tindal, 2013). Some tasks require only the detection and transcription of explicitly stated information in the text, and the while the others demand the combined use of text information and the previous knowledge of the reader.

Several reading comprehension taxonomies have been formulated in an attempt to categorize the different demands of reading comprehension tasks (Barret, 1976; Catala, Catala, Molina, & Monclus, 2001). These taxonomies seem to converge in four domains designated by Catala colleagues (2001) as literal comprehension (LC), inferential comprehension (IC), reorganization (R), and critical comprehension (CC).

Literal comprehension (LC) entails the recognition of information explicitly stated in the reading section. Inferential comprehension (IC) emerges the reader's prior knowledge is activated, and expectations and assumptions about the text contents are made based on clues provided by the reading. Reorganization (R) implies a new way of organizing information through synthesis, schemes or summaries. Critical comprehension (CC) includes making judgments with subjective answers, relating to the characters or the author's language and personal interpretations. This conceptualization can allow teachers to design specific instructional activities to promote each type of comprehension. Tests should reflect those different types of comprehension in order to provide relevant feedback on student's feedback (Basaraba et al., 2013). Thus, the reading comprehension section contains the reading passages and questions about the passages. The questions are about information that is stated or implied in the passage and about some of the specific words in the passages.

In addition, reading texts refers to a process of extracting and understanding the required information from publications (Grellet, 1981 as cited in Biruk, 2013). It is also described by Wallace (2003) as a reaction to written texts by a reader in order to communicate with the writer. Similarly, reading a text as "process[ing] text meaning through some process of interaction with print" Alderson (2003, as cited in Bruk, 2013). These definitions tell us that reading is an active process where the reader deals with the written materials by guessing, predicting, analyzing and interpreting the message which is beyond mere deciphering in relation to his/her background knowledge. Reading comprehension tests therefore, should be well organized and allow students to facilitate their test performance rather than allowing its absence to inhibit performance on test-takers' ability (Alderson, 2000).

2.2. Reading and Comprehension

Reading comprehension is a complex undertaking that involves many levels of processing. The fundamental aspects of comprehension is the ability to deal with unfamiliar words encountered in text, the main idea, detailed information, and inferences (Paynter, et.al., 2005).The two terms are inter-related with each other. Besides this, comprehending is to understand and to attain the level we need to read. In relation to this, Gray (1960, as cited in Biruk, 2013) tells “.... during processing leads to the contraction of understanding or comprehension. As we read constantly, comprehending strategies, making connections and inferences, summarizing and synthesizing information, and perhaps even critiquing or analyzing the text in some ways.” In contrast, people often speak of reading and comprehending as two different or connected processes. As Grellet, (1981, as cited in Biruk, 2013), said “Well, he can read it, but he does not understand it. I take the position reading comprehends without understanding, a person may be noticing and responding to graphic symbols but not processing them in a meaningful way that is required of readers.”

2.3. The Importance of Reading

Reading is one of the four language skills and serves as an instrument for communication of thoughts from one person to another, and it is a means to gain several things. For instance, by reading a newspaper, magazines, and other sources one can learn about the current social, economic and political situation of the world and by reading comprehension test questions the test-takers saw their test performance.

According to Cross and Carney (1950), “Much of the success of the students in school and in later life depends upon how well they can read, and how well they can use the meanings which they have gained from their reading situation.” Reading is needed in both classroom and real life situations. According to Bright and Mc Gregory (1972, as cited in Biruk, 2013), it is the major language skill that enables learners to understand ideas, facts, concepts and thoughts.

Scholars have suggested that reading has two purposes. First, if a person has an effective reading ability, he will succeed in his academic performance; secondly, reading ability also helps one for good social life with the society (Bright, 1970 as cited in Biruk, 2013). Reading skill for students contribute a lot not only in their preparatory school level but also it helps in the university level. In relation to this, further education depends on quantity and quality of reading for academic purposes for assessment using classroom tests.

2.4. What is Test?

According to Brown (2004), a test is defined in the following way:

A test, in simple terms, is a method of measuring a person's ability, knowledge, or performance in a given domain. To qualify as a test, the method must be explicit and structured.

Next, a test measures performance, but the results imply the test-takers ability. Most language tests measure one's ability to perform language, that is, to speak, write, read, or listen to a subset of language (p.14).

In addition to this, Carroll (1968) provides the following definition of a test:

A psychological or educational test is a procedure designed to elicit certain behavior from which one can make inferences about certain characteristics of an individual. From this definition, it follows that a test is a measurement instrument designed to elicit a specific sample of an individual's behavior. As one type of measurement, a test necessarily quantifies characteristics of individuals according to explicit procedures. This definition reflects the primary justification for the use of language tests and has implications for how we design, develop and use them.

However, it is precisely because any given sample of language will not necessarily enable test user to make inferences about a given ability that we need language tests. That is, the inferences and uses we make of language test scores depend upon the sample of language use obtained. Language tests can thus provide the means for more carefully focusing on the specific language abilities that are of interest (p.68).

In addition, tests are prepared administrative procedures that occur at identifiable times in curriculum when learners master all their faculties to offer peak performance, knowing that their responses are being measured and evaluated (Brown, 2004).

2.5. The Purpose of Tests

Test is given formally while a course is on progress and serves as instrument for collecting information. Its purpose is to assess learning progress, identify if there is learning difficulties (Demp, 1994). Again, Bachman (1990) states two major uses of language tests are serves as a cradle of information for making decisions within the context of educational programs and as indicators of abilities or attributes that are of interest in research on language, language acquisition, and language teaching.

Teachers are usually expected to use two primary techniques to gather information about their pupils' achievement, instruction, and classroom atmosphere. The first technique involves the use of paper-and-pencil tests, and the second involves the use of observational techniques (Thorndike, et al., 1977; Gronland, 1981).

A standardized test is any test with uniform questions that is administered and scored in a consistent manner to ensure legal defensibility. It is also often used in education, professional certification, psychology, the military and many other fields. It also widely used, fixed in terms of scope, difficulty and format, are usually significant in consequences (Thissen & Wainer, 2001). It is usually held on fixed dates as determined

by the test-developer, educational institution, or governing body, which may or may not be administered by the instructor, held within the classroom, or constrained by the classroom period.

According to Thissen and Wainer (2001), a non-standardized test is usually flexible in scope and format, variable in difficulty and significance. Since these tests are usually developed by individual instructors, the format and difficulty of these tests may not be widely adopted or used by other instructors or institutions. It may also be used to determine the proficiency level of students, to motivate students to study, and to provide feedback to students. In some instances, a teacher may develop non-standardized tests that resemble standardized tests in scope, format, and difficulty for the purpose of preparing their students an upcoming standardized test. Finally, the frequency and setting by which a non-standardized tests administered are highly variable and are usually constrained by the duration of the class period.

A test may be developed and administered by an instructor, a clinician, a governing body, or a test provider or researcher. As with the development and administration of educational tests, the format and level of difficulty of the tests themselves are highly variable and there is no general consensus or invariable standard for test formats and difficulty. Often, the format and difficulty of the test is dependent upon the educational philosophy of the instructor, the subject matter, class size, policy of the educational institution, and the requirements of accreditation or governing bodies (Thissen & Wainer, 2001).

With the issue of test preparation, from the perspective of test developer, there is a great variability with respect to time and effort needed to prepare a test. Likewise, from the perspective of a test-taker, there is also a great variability with respect to the time and needed to obtain a desired grade or score on any given test. When a test developer constructs a test, the amount of time and effort is dependent upon the significance of the test itself, the proficiency of the test-taker, the format of the test, class size, deadline of the test, and the experience of the test-developer. Thus, the process of test construction

has been greatly aided in several ways. For one, many test-developers were themselves students at one time, and therefore, are able to modify or outright, adopt test questions from their previous tests (Thissen & Wainer, 2001). In general, tests developed and administered by the individual instructors are non-standardized whereas tests developed by testing organizations are standardized.

Paper-and-pencil tests can be either standardized or non-standardized tests. The following table is a summary of their characteristics.

	Paper-and-Pencil Tests	
	Non-standardized tests	Standardized tests
Purpose	Assessing learning outcomes intended for one time use in a class	Assessing general ability, attitudes and learning outcomes throughout all levels of school.
Examinees and test taking situations	A single group of pupils at a single point in a time	Different groups of pupils at different time and place
Procedure	No definite administering, scoring, and interpreting	Administered, scored, and interpreted in the same way across different groups for comparison
Prepared by	Mostly by school teachers	Companies or organization consulting test specialists, curriculum experts and skilled teachers
Example	Mostly classroom tests, unpublished tests	Mostly published tests such as Scholastic-aptitude tests, Intelligence tests, Achievement tests, National exams, etc.
Item Format	Mostly true-false, matching, supply, multiple choice items	Mostly multiple choice, pictorial items

Source: ICDR, 2002, p.201

One of the item formats in paper-and-pencil tests prepared in schools is supply tests (ICDR, 2002). The following table shows the characteristics of the item format of supply tests.

	Supply Tests	
	Short-answer and completion	Essay Tests and Paragraphs
Format	. Question form(short answer) . Incomplete statement (completion)	. Answering with limited number of lines or words (restricted essay or paragraph)
Behavior assessed	. Simple learning outcomes	. Complex learning outcomes (i.e., organizing, synthesizing ideas, use of language etc.)
Scoring	.Relatively subjective and difficult	. Subjective and unreliable
Advantage	. The easiest to construct . Avoid guessing . Pupils don't have to write out answers	. Measure areas which cannot be measured by other ways . Avoid guessing
Limitation	. Do not measure complex outcomes . Scoring is subjective and difficult	. Scoring is subjective and unreliable

Source: ICDR, 2002, p.203

2.6. Construct-Irrelevant Factors

Construct-irrelevant factors are the factors that cause variation on test-takers test performance which represents issues beyond the content and/or skills that are the focus of the test (Fairbairn, 2007). According to Kucuk and Walter (2009), the test structure such as the vague instructions of the test, the layout of the exam paper and its legibility, the

issue of stating the time given on the exam paper; and testing conditions such as distracting sounds and noises, the amount of light in the classroom, the amount of air in the classroom, time allocation, way of instructor or invigilator support of test-takers during the test, insufficient test materials such as answer sheet without reserves and ample spaces on which the test-taker can fill in answers and the degree of temperature hindered students' performance in tests.

The most important features of tests are its validity, and in rough terms, a test is said to be valid if it measures what it claims to measure. That is, when a test is valid, the construct (ability, skill, trait or domain of knowledge) that it is designed to measure is the source of examinees' scores on test. For instance, if the items on reading comprehension test were printed in a very small type, examinee's scores would vary due to the variability in visual acuity and the test would be measuring visual acuity rather than just the reading comprehension. This problem is called the construct-irrelevant factors (Haladyna & Downing, 2004). In addition, they stated that testing students during the typical lunch or playground, holidays or exciting events, not informing students about the test when and where the test is given without their consent is a source for anxiety, and inappropriate instruction are some of the factors that affect test-takers test performance. Therefore, the construct-irrelevant factors affect students' language test-performance beyond the skill, knowledge, and ability that the test is intended to measure.

According to the Council of Europe (2011) some source of construct-irrelevant factors that affect students' performance on tests includes the day of the test session (the weather, the administration, etc., may be different) as a result, the individual test-takers may perform better or worse on a given day; the markers may perform or mark differently; size (inability to accommodate the numbers required with sufficient spacing between desks); the inappropriate room shape (that does not allow invigilators to see all test-takers clearly to control cheating); insufficient facilities (test materials); the method of delivery that is not both timely and secure so that all materials are not in place by the time of administration.

2.6.1. The Problems to Language Test Performance in Reading Comprehension Test

A common fault of tests written for the students of particular teaching institutions is the supposition that the students all know what is intended by carelessly worded or written tests (Hughes, 1989). The frequency of the complaint that the students are unintelligent, have been not good, have willfully misunderstood what they were asked to do, reveals that the supposition is often unwarranted (Bachman, 1990). Test writers and developers should not only rely on the students' powers of telepathy to elicit the desired behavior (Hughes, 1989).

Studies have unanimously indicated that the observed performance of a test taker is a representation of his abilities, test method facets and other factors that affect test performance (Bachman & Palmer, 1986; Alderson et al., 1995). Also, reading difficulty, ambiguity, vague, long and complex sentences, difficult vocabularies are some of the extraneous factors that affect language test performance (Thorndik, et al., 1977; Ebel, 1979; Gronland, 1981). These extraneous factors are the challenges that affect students' performance on language tests particularly reading comprehension test beyond the ability, skill and knowledge that the test is intended to measure.

Every test is designed to assess one or more construct. According to Chappelle (1994, as cited in Buck, 2001) construct is defined through two ways. One is to define competence (knowledge, skills, and abilities) expected to demonstrate by the test-taker. The second is to define the tasks expected to be performed by the test-taker. In addition, constructs are a systematically organized concept, which is drawn from a theory of ability to be tested. They are the components of a theory and the relationship among them is determined by a theory. They are also realized through tasks to be performed by test-takers (Brown, 2004; Hughes, 2003; Alderson et al., 1995).

According to Brown (2004), the development of reading skills is a complex process and the constructs of reading depend on a model of reading and myriad factors influencing reading. There is a range of constructs to be assessed in reading a second or a foreign

language and the tester needs to define the construct to focus upon relevant abilities being measured and considers the process of assessment.

The constructs of reading come from a theory and can be incorporated into test specifications which describe test tasks intended to measure skills, sub-skills and strategies of reading. Some theories of reading explain the presence of different constructs such as skimming, scanning, reading in details and etc. And these are different from another (Alderson, 1995). Another example of skills to be tested in reading may be the skills of synthesis and evaluation. As elaborated by Alderson (2000, p.118) “Such a construct may be more narrowly defined as the ability to identify, distinguish and compare evaluate evidence and opinion.”

Apart from this, students’ test performance on reading comprehension test is influenced by the construct-irrelevant challenges that influences students test performance beyond the skill, ability and knowledge that the test is intended to measure (Abedi, 2002).

2.6.2. The Factors Influencing Students’ Performance in English Classroom Tests

Students’ test-performances on structured and non-structured classroom tests are often low due to a number of issues beyond the knowledge, skills, and abilities that the test is intended to measure (Abedi, 2002). Fairbairn (2007, p.4) states that “Matters of language, ineffective use of test accommodations, and unfamiliarity with the test and/or question format may hinder student performance by introducing and/or falling short of eliminating construct-irrelevant factors.”

According to Bachman (1996), badly typed and/or too much text in too a small space and inappropriate time allocation causes anxiety and lowers students’ test achievement in their test result. Therefore, all the above mentioned factors may impede test performances in such a way that the performances represent issues beyond the content and /or skills that are the focus of test.

2.6.2.1. The Factors Related to Language

Published educational literature suggests that there are three types of language. These are everyday language, academic language, and test language. The inability to use test language during language test when but vocabulary is not tested is a challenge for students' performance on tests that in turn affects students' test scores/or results (Kopriva, 2000). He also suggests the absence of "plain language" in test development on the side of test developers that reduce the proper students' test performance. In reading comprehension tests, for example, if the language of presentations is harder to understand than the passage itself, the reader is presented with an additional layer of difficulty and we cannot tell whether poor performance is due to the passage difficulty or to that of the questions (Alderson, 2000).

Construct-irrelevant factors that are related to language and affects test-takers performance are unfamiliar vocabulary that are not related to the construct being measured; cultural references or idiomatic expressions that are not equally familiar to all students; syntax that may be confusing or ambiguous such as negatives or double negatives; low frequency, long or morphologically complex words and long sentences; sentence structure that may be confusing or difficult to follow such as passive voice or sentences with multiple clauses; syntax or vocabulary that are above the test's target grade level (Abedi, 2006).

2.6.2.2. The Factors Related to the Ineffective Use of Test Accommodations

Hacuta (2009, p.25) states that "test-accommodation is any change to standardized testing conditions intended to make the test more fair and accessible for an individual or sub-group that does not change the construct being measured." These changes may include, but are not limited, to change in the presentation of the assessment, the environment in which the assessment is administered, time allowed for assessment, or additional materials or equipment to be used by students during the assessment.

Variations in administrations such as too little amount of time for tests that need much time; the activities of invigilator such as the disruptive behavior while the test-takers doing the test; testing students in a large-group setting without conducive seating arrangements in a narrow classes affects students' performance on tests (Fairbairn, 2007).

2.6.2.2.1. Two Level Taxonomy for Organizing and Analyzing Test-accommodations

According to Rivera, Collumn, Willner and Sia (2006) accommodations can be divided into two for organizing and analyzing test accommodations for test-takers. This taxonomy divides accommodations into those that offer direct linguistic support (in the native language and English) and indirect linguistic support to students. In relation to the extent to which they offer linguistic support, Hacuta (2009) states that the direct linguistic support includes the use of L1 or L2 dictionaries and glossaries, and oral administration of directions in L1 or L2; and the indirect linguistic support includes extended time, additional breaks and modifications to the test setting. The inappropriate usage and shortage of using such accommodations may affect students test performance.

2.6.2.3. Students' Unfamiliarity with the Test and/or Question Format

Familiarizing students with the format of the of test questions is not "Cheating "rather; it is creating the possibility for test scores to more accurately reflect students' performance (Hacuta, 2009). If students answer items incorrectly because of confusion about the test question format, their performance fall short of indicating to stalk holders what students know and are able to know and are able to do within that content domain (Fairbairn, 2007).

2.6.2.4. The Factors Related to Badly Typed Texts and Test-Questions

Badly typed and too much text or test questions in too a small space and inappropriate time allocation causes anxiety and will lower students' test performance which in turn affects the test result beyond the ability that the test is intended to measure (Hughes,1989).

2.7. Strategies to Alleviate the Problems of Construct-irrelevant Factors

The following sections address the strategies that are specific to the areas of language, accommodations, test and/or item format, and badly typed and the usage of too much text into a small space texts or questions that will enable test developers and users to elicit more accurate representations of what students know and can do.

2.7.1. Language

Gardon and Hanauer (1995, p.302) remarked, “Having questions written in plain language would facilitate the test-takers’ understanding of what is being asked in a particular item. This would decrease the chances of the test-taker providing the wrong information due to misunderstanding the comprehension questions.”

Abedi and his colleagues have done a great deal of work in the area of linguistic simplification of test items and provide examples of ways to implement this strategy (Abedi, Courney, Mirocha, Leon, & Goldberg, 2005; Abedi, Lord, Hofstetter, & Baker, 2000; Abedi, Lord, & Plummer, 1997). Kopriva (2000) also addresses the notion of “plain language” test development, providing guidance in how to implement this strategy. Those who develop tests for ELLs, whether standardized or classroom based, are urged to incorporate these kinds of techniques for reducing language related construct-irrelevant variants in tests. Specific strategies drawn from the research literature include the following:

2.7.1.1. Use Simple Grammar and Sentence Structure

Ensure that the language of tests is clear and easily understand. For example, sentences in present tense are likely to be more readily accessible to ELLs than those in other tenses and shorter sentences are easier to digest than longer ones (Abedi, 2002).

Hacuta (2009) presented the proper usage of language in the following ways:

Using clear and accessible language is a key component of minimizing construct-irrelevant variances. However, do not simplify language test that is part of the construct being assessed (e.g., the passages on a reading comprehension test or challenging vocabulary that is part of the construct of a subject area test). In other cases, though, the language of presentation should be as simple and clear as possible. Some of the general guidelines for using accessible language are the use of vocabulary that will be widely accessible to students. In addition, avoid colloquial and idiomatic expressions, words with multiple meanings, and unduly challenging words that are not part of the construct; and keep sentence structure as simple as possible to express the intended meaning. For Second Language English Learners, a number of simple sentences are often more accessible than a single more complex sentences (p.14).

2.7.1.2. Use Common Vocabulary Wherever Possible

When vocabulary is not being tested, test items can be more on accessible through the use of everyday vocabulary rather than lesson common terms (Kopriva, 2000). Language used on tests is different from the everyday language that most ELLs quickly learn and even from the academic language used in instructional settings. Published educational literature suggests that this test language is a third type of language (in addition to everyday and academic language) that students must acquire in order to be successful on tests (Calkins, Montgomery & Santman, 1998; Stevens, Butler, & Castellon- Wellington, 2000).

2.7.2. Test Accommodations

According to Hacuta (2009), test accommodations are a means of minimizing difficulties that students face in testing situations that relate specifically to their status as L₂. Ideally

an accommodation should be one that benefits students as second language learners but not native speakers of English.

The main purpose for providing accommodations to examinees is to promote equity in assessment. That is it serves as a means to ensure that ELL have the same opportunity as other students to demonstrate their performances and promote equity for all students in test taking situations and activities. The testing accommodation includes testing in smaller, rather than larger groups, extended or balanced testing time and extra-breaks during testing (Abedi, 2006).

Some researchers outline various accommodations that are appropriate for second language learners. Fairbairn (2007) states common accommodations include variations in administrations, such as appropriate time or testing students in a small-group setting, and in a response mode are the conducive test accommodations. The following is a list of considerations for using test accommodations:

2.7.2.1. Utilize Accommodations that do not affect what is being measured

It would be inappropriate and unacceptable to read a reading test aloud to a student if that test is designed to measure the students' ability to comprehend text that she/he reads silently. In this case, the "accommodations" has a direct affection on what was being assessed or measured such that the student's scores would not reflect the intended ability (Hacuta, 2009). Since the goal of testing is to estimate students' knowledge, skills, and/or abilities and performance in specific areas, accommodations that affect what is being measured negatively impact the value of testing procedures and the resulting scores.

2.7.2.2. Use Accommodations that are Students Responsive

Rivera, Collum, Willer, and Sia (2006) state that accommodations are intended to minimize the cognitive resources students or test-takers need to process the language of the test and maximize the cognitive resources available for accessing the context of the

test. In addition, SLA research reveals that repetition and clarification represent potentially beneficial accommodations.

2.7.3. Students' Unfamiliarity with the Test and / or Question Format

Students' unfamiliarity with the test and / or question format is a third issue deserving of attention when testing students. Different cultures have different ways of expressing and evaluating knowledge and skills (Hacuta, 2009). For example, multiple choice testing is common in one countries context; other forms of assessment are more in other contexts. Students (ELLs) who are unfamiliar with the format of the test itself or the test questions may likely have difficulty in demonstrating what they know and are able to do on such tests (Fairbairn,2007). Therefore, Popham (2005) clarifies that general test taking practice is perfectly acceptable, as is familiarizing students with a variety of test and question formats. A number of test preparation strategies recommended by a variety of researchers that can be applied in both classroom based and large scale testing contexts are the following.

2.7.3.1. Align Testing with Instructional Practice

In classroom based tests, students are most likely to be able to best demonstrate what they know and can do if the test takers are familiar to them (Popham, 2005). Lack of test task familiarity may introduce construct-irrelevant variance similar to that resulting from inappropriate language demands on tests. Teachers are urged to make use of test/item formats that are familiar to students based on their classroom experiences.

2.7.3.2. Teach Students How to Negotiate Different Item/Test Formats

Teachers can model the test taking process and discuss student work on various types of tests, resulting in a more accurate demonstration of what student know and do (Hacuta, 2009). In addition, giving students time to reflect on and share their ideas about how to “attack” various kinds of test-questions can prove very helpful.

2.7.3.3. Ensure that Students Know the “Rules of the Testing Game”

Clarifying for the students’ the rules of test-taking activities and penalties for breaking the rules whether it is acceptable to go back to change answers on finished test sections, and whether students can write in their test booklets or use scratch paper (Fairbairn, 2007).

2.7.3.4. Allow Students to Experience the Testing Conditions

The time constraints that students will face during the actual test, the seating arrangements (especially if they are different from the usual classroom set-up), and the materials to be used during testing such as answer sheets and others serve as a strategy to overcome the test-takers unfamiliarity with the test format (Hacuta, 2009).

Students may likely need practice in budgeting their time, which could take the form of timed classroom activities throughout the year (Abedi, 2002). In addition, techniques for coping with stressful situations (such as taking a deep breath or stretching) can be brainstormed and used by students. Test-takers who are tempted to spend an inordinate amount of time “bubbling” in their answers perfectly may well run out of time, resulting in scores that do not truly reflect their ability. Likewise, practice in using support materials or technology will assist students in using those tools effectively during testing (Fairbairn, 2007).

2.7.3.5. Teach Specific Test Taking Techniques

There are strategies that assist students in demonstrating what they know on classroom achievement tests. According to Fairbairn(2007), the techniques include“ strategies such as surveying the entire test to get an idea of what is included; reading the questions before reading the passage; using the “process of elimination”, and budget their time appropriately.”

2.7.4. The Badly Typed and Usage of Too Much Text in Too a Small Space

Ensure that tests are well laid out and perfectly legible. Test developers should be aware of formatting issues. Fonts, font sizes, and line breaks in paragraphs should all receive careful reviews (Hacuta, 2009). Too often, institutional tests are badly typed (or handwritten), have too much text in too small a space and are poorly reproduced, As a result, students are faced with additional tasks which are not once meant to measure their language ability. Their variable performance on the unwanted tasks will lower students' score on a test (Hughes, 1989).

CHAPTER THREE

3. RESEARCH DESIGN AND METHODOLOGY

This chapter presents the research design, study population, sampling of the study, data collection instruments, data collection procedures, data analysis procedures and ethical consideration.

3.1. Design of the Study

A descriptive research design was employed to assess the construct-irrelevant factors influencing students' test performance in a reading English comprehension test. To achieve the intended objectives of the study, Mixed Research Approach in which the researcher incorporates qualitative and quantitative methods to answer the research questions was used. It helps the researcher cross-check the data obtained and get comprehensive information in studying the proposed problem other than using either qualitative or quantitative method alone (Onwuegbazie & Leech, 2006). In reference to this, Sharma (2000, as cited in Alemayehu, 2008) states that a descriptive survey is useful to identify the present condition and point to present needs. And also it is helpful in establishing the existence of phenomenon by describing explicitly the phenomenon under investigation.

The researcher used quantitative method of data gathering tools such as questionnaire and model open-ended reading English comprehension test and at the same time interview and document analysis as a means of data collection system which fell in qualitative method of research design. Such triangulation has a complex phenomenon that requires elucidation, multiple methods that/which are suitable where some aspects of data need to be evaluated more fully.

3.2. Participants of the Study

To conduct this study, the researcher used Grade 11 students of Ano and Sire Preparatory Schools. The reason for focusing on this grade level was that the researcher is working in Ano Preparatory School, and Sire Preparatory School is the nearby school to the researcher's work place. This enabled the researcher easily to have access to the schools and to obtain data. Moreover, the researcher believed that the students at this grade level can hopefully understand the questionnaire set to be completed by them when compared to Grades 9 and 10 students. But Grade 12 students were excluded because they were not conducive for the researcher to gather data from them for they were preparing themselves to take the national exam.

The population size is three classes with about 90 students which are 24, 25 and 41 respectively in each section of the classroom registered in 2007 E.C. at Ano Preparatory School. Nine students were selected by systematic random sampling and they participated in answering questionnaires and doing the open-ended reading comprehension test. Additionally, the population size of Sire Preparatory School is seven classes with about 352 students which are 43, 43, 43, 43, 60, 60 and 60 students respectively in each section of the classroom registered in 2007 E.C.. Thirty-five students were selected by using systematic random sampling to answer the questionnaire and do the short-answer open-ended reading comprehensions test.

The participant of the study also included all English teachers who are teaching Grade 11 students. That is, two teachers from Ano and three teachers from Sire Preparatory schools were included by purposive sampling. The reason for using English teachers was, because they were probably familiar with the construct-irrelevant factors that affect students' performance on a reading comprehension test as they were test-developers in the English classroom tests.

3.3. Sample Size and Sampling Techniques

The study was conducted in Western Ethiopia, Oromia Regional State, and Eastern Wollega Zone. There are 17 Woreda's in this zone. Out of these woreda's, Sire and Ano, Preparatory schools were selected by purposive sampling. It is the nature of data gathering instruments such as administering open-ended reading comprehension test and document analysis that restricted the researcher to use only two schools from 17 woreda's found in this zone because it requires too much time and resources. In relation to this, the main factor considered in determining the sample size is the need to keep it manageable enough (Neuman, 2003). Also this enables the researcher to drive detailed data at affordable cost in terms of time, finances and human resources (Mugenda & Mugenda, 2003). Therefore, due to its convenience for the researcher, Grade 11 students of these schools were selected because they are found in the nearby woreda's to the researcher.

One of the methods used in sample size determination for a given study is the type of methods used in that study. This study used both qualitative and quantitative research methods. In relation to this, if the nature of the study is qualitative, small sample is sufficient but if the study is quantitative, a large numbers of samples is demanded (ICDR, 2002). Thus, in a mixed research design, for the sake of accuracy in a relatively small total population, a representative sample of 10% is sufficient.

Gay and Airasian (2000, as cited in Best & Khan, 2003) argued, it is fair and appropriate to use 10% of the total population if the population is less than or equal to 1000. Therefore, the total population of Grade 11 students of Ano and Sire Preparatory schools is 442 which is less than 1000. In addition, Neuman (2003) indicated that 10% - 20% is an adequate sample in a descriptive study. Thus, out of 442 students 44 students were drawn as a sample of the study. Nine students from Ano, thirty-five students from Sire Preparatory schools were selected by systematic way using the systematic fraction method, which is symbolically represented by K . Employing this method samples are drawn to be equal to N/n , where "N" is the total number of units in the target population, and "n" is the number of units of the sample (ICDR, 2002). Therefore, $N=90$, $n=9$, the

sampling fraction is 9; N=352, n=35, the sampling fraction is 35; N=90, n=9, the sampling fraction is 9 in Ano, and 35 in Sire respectively. The number selected randomly will have the n^{th} term. Finally, the codes on the sampling frame that correspond to the number drawn were identified and included in the study.

This study was then used both quantitative and qualitative research methods. Thus, the researcher used purposive sampling in selecting the two schools. For the quantitative research method, 44 students were selected using systematic random sampling and they answered the questionnaire and did the sample reading comprehension test which was prepared for this purpose. For the qualitative research method the researcher used all the 5 teachers who are teaching Grade 11 students by using purposive sampling to get their opinion by semi-structured interview. This could help the researcher to answer the questions raised previously in the research questions.

The systematic random sampling helped the researcher to give equal chance for the targeted population and to get relevant information for the study from the students answered the questionnaire and did the sample reading comprehension test. In addition, purposive sampling method helped the researcher to get information about the problems related to these construct-irrelevant factors to reading comprehension test from teachers' responses to the interview.

3.4. Data Collection Instruments

Researching and examining tangible evidence requires intensive involvement of the researcher and the researched in the given study (Sarantakos, 1993, as cited in Biruk, 2013). In relation to this, this study employed reading comprehension test prepared for this purpose to check the construct-irrelevant factors that affect students' performance in doing this test. Thus, it was a major data gathering tools and questionnaire and interview for students to check their judgments about their test performance in an open-ended reading comprehension test. English teachers' interview served as a major tool to get a direct insight into real practices concerning the assessment of students' performance on

test beyond the skill, ability, and knowledge that the test is intended to measure. Thus, according to their relevance to this study, the major principal tools such as a model reading comprehension test, checklist questionnaire, document analysis, and semi-structured interview are going to be discussed as follows.

3.4.1. Test

The instrument of this study is short-answer item reading comprehension test. This reading comprehension test comprises four essay paragraphs and each of the paragraphs is from different modes of writing namely, descriptive, expository, narrative and argumentative essays respectively.(See Appendix C and D). This is because the researcher believed that different essays may have different influences on the test performance of students. Thus, to balance the chance of student's performance on doing the test the researcher used the above four essays.

To answer the research questions that are related to the problems of language factors, ten questions for the first session which believed to contain unfamiliar languages was presented to the student respondents and the questions were focused on the main idea, and recognize specific details. But understanding vocabulary from the text and recognizing inference were not contained in the questions because of the nature of the research questions to be answered by this study. Again with the same version of the given paragraphs, the same questions but questions with familiar languages were presented for the student respondents to see the disparity of language factors on the students' test score.

The student respondents took short-answer item in reading comprehension test twice. This is because the researcher believed that when the first session of a test was presented with unfamiliar languages and the second sessions of the test was presented with familiar languages; it helps to see the construct-irrelevant factors of language on students' test performance. According to Alderson et, al. (2000), the reliability of objective tests is measured by test-retest, parallel form, split-half and etc. but the reliability of subjective tests are measured by inter-rater reliability. Therefore, since open-ended questions require

subjective scoring, the researcher used two scorers from the two schools to see the reliability of the test in scoring the short-answer open-ended questions in reading comprehension test. Thus, open-ended questions require subjective scoring, in this case the reliability of subjective tests is measured by calculating the reliability of the marking; this is done by the inter-rater reliability (Kappa) which refers to the degree of similarity between different examiners: can two or more examiners, without influencing one another, given the same marks to the same test of scripts (Wang, 2009).

Finally, the researcher used Kappa which is used for assessing the inter-rater reliability and has a range from 0-1.00, with larger value indicating better reliability. Thus, an inter-rater reliability analysis using the Kappa statistics was performed to determine consistency among raters (Wang, 2009). Specifically, a more complete list of how Kappa might be interpreted is given in the following table (Landis & Koch, 1977).

Kappa	Interpretation
< 0	Poor agreement
0.0-0.20	Slight agreement
0.21-0.40	Fair agreement
0.61-0.80	Substantial agreement
0.81-1.00	Almost perfect agreement

According to Landis and Koch (1977) Kappa is not an inferential statistical test, and so there is no HO. Therefore, the researcher used the KAPPA.SAV in SPSS version 20 to estimate inter-rater reliability. In addition, the pilot test was conducted to check the validity of the test with 22 students. These students were not included within 44 student respondents because to take care for the corruption of data resulted due to respondents' awareness about the areas of data assessed by the test that in turn affects the validity and reliability of the test result.

Therefore, to analyze data obtained from test-scores the researcher used the Sturge's Rule. According to Jimma University CDE (2010) it is possible to fix the number of

classes or intervals (K) using Sturge's Rule method, which is a personal judgment depending upon the nature of investigation one can decide with the help of the rule stating that:

$$\begin{aligned}\text{Number of classes (K)} &= 1+3.322 \times \log N \\ &= 1+3.322 \times \log 44 \\ &= 6.459 \\ &\sim 7\end{aligned}$$

Where N= total number of observations and log is common logarithm.

In relation to this, if the number of classes is known and if it is decided to use classes of equal size, the determination of the size is relatively simple. Since the class size depends upon the number of classes and the extent to which the values of the variable are spread or dispersed, the following sample formulae can be used. Thus, the class width can be calculated in the form of:

$$\text{Class width} = \frac{\text{Range}}{\text{Number of classes}} \quad \text{or} \quad Cw = \frac{R}{k}$$

Where R= Range= Highest value – Smallest value

The relative frequency of a class shows the relative concentration of items in a given class interval to the other classes of a frequency distribution. Thus, the relative frequency of a class= $\frac{\text{class frequency}}{\text{total frequency}}$

Finally, to see the difference between the familiar and unfamiliar test results, the paired sample test was used to compute the mean differences. According to Park (2009), the paired sample T-test is used to compare the two related means. It is also used to compare two population means when there are two samples in which observation in one sample can be paired with observations in other sample or the before and the after observation on the same subject.

3.4.2. Questionnaire

The questionnaire for the students was used as one of the data collecting instruments for this study. It was used because it is easy to administer to a number of participants in one place at a time and it also enables the researcher to obtain data from large group of individuals within a short period of time.

It is planned as a major data collecting tools for this study because the researcher thought it helpful to get necessary information about students' judgments about the challenges they encountered on their performance on the test beyond the ability, skill, and knowledge that the test is intended to measure.

According to MOE (2002), "In constructing questions for a classroom test and constructing questions for a questionnaire the researcher will be concerned about the content, format and language of the question." Therefore, content questions for background information of the respondents and question format which is the common techniques of structuring the response categories of closed-ended questions and the language usage to convey clear message in the questionnaire would be taken in to consideration. Thus, based on the nature of this study, from the question format, a check list was employed to gather data from students. In a check list questionnaire, the respondent is required to give "Yes" or "No" response and then provided their suggestions. (See Appendix A). Additionally, to check the validity of the questionnaire, pilot test was conducted with 22 students. Thus, questions number 4 and 10 were corrected based on the students' suggestions. The former question number 4 was stated as "Is the vocabulary in context asked for its meaning written in bold letters?" and that of question number 10 was stated as "Are the test questions provided with difficult vocabularies?" Therefore, students were confused to respond these two questions under "Yes" or "No" and leave the space as it was and suggested the addition of the word "sometimes" for question number 4 and "always" for question number 10.(See Appendix A).

The checklist questionnaire was presented with Afan Oromo because the student respondents can fully understand the questionnaire set to be filled by them when compared to English language questionnaire. The open-ended questions that require students' suggestions provided simultaneously with "yes" or "no" to be answered by the student respondents and then thematically organized by the researcher according to the research questions of the study to be answered. Thus, the questionnaire was distributed to student respondents two days after the mid-exam given in their respective schools to keep the reliability and validity of the data obtained through the test.

3.4.3. Semi- Structured Interview

Robinson (2000, as cited in Muhammad, 2007) states that semi-structured face-to-face interview is mostly chosen as it is important, because it enable the researcher to access the degree of the interviewees' interest and participation. Furthermore, it is also essential as it is flexible to balance by structure and the quality of the data so obtained (Gillham, 2005 as cited in Dessalegn, 2014). It involves a series of open ended questions based on the topic areas the researcher wants to cover and provides the interviewer to use clues or prompts to encourage the interviewee to consider the question further. In addition, it involves some sort of flexibility so that other important information is added gradually.

Therefore, all 5 English teachers who are teaching Grade 11 students were included in face-to-face semi-structured interview which contains six questions regarding the construct-irrelevant factors, the adverse effects and strategies to overcome the construct the construct-irrelevant factors that affect students' test performance in reading comprehension test. The interview which contains six questions was presented in the English language because the researcher thought that the teachers understood the English language. Their responses were recorded by the tape recorder for the later careful data analysis. The whole question item was open ended. The researcher used the interview to triangulate the data that was obtained through the questionnaire and model reading comprehension test and besides it helped the researcher obtain deeper insights into the research problem.

3.4.4. Document Analysis

Document analysis refers to the analysis of any written or recorded material not prepared for the purpose of the research or at the request of the inquirer. Public documents, archival records, personal documents, and formal studies and reports are the major types of documents (ICDR, 2002). Exam questions are an example of archival records. It helps the researcher when documentary review is useful and the research questions may be answered by the existing documents and to study the past events. Therefore, the researcher used reading comprehension exam questions from the archive of exam-bank prepared by English school teachers in 2014/15 at Sire and Ano Preparatory schools for mid-exam and final-exam to answer the construct-irrelevant factors that affect students' performance in reading comprehension test particularly the factors related to language, format of the text (passages) and test item and its written visibility. This teacher made tests helped the researcher to tailor the specific research question. Therefore, 5 questions were prepared to analyze the exams prepared by teachers at Ano and Sire preparatory schools in 2014/5 academic year only to cross-check the data obtained from students checklist questionnaire and finally the data collected was thematically organized.

3.5. Data Collection Procedures

The data for the study collected through the following steps. Before administering the questionnaire the researcher explained the purpose of the study for both the school principal and the subjects of the study. Soon after, the participants of the study were asked to fill the check-list questionnaire after their school mid-exam. The model test for the study was then given in the second session.

Finally, all the five English teachers were interviewed and the data was collected. Having completed the process of collecting data, the researcher analyzed the data which was obtained using the aforementioned instruments.

3.6. Data Analysis Procedures

The Mixed Research Design that is used in this study required both the quantitative and qualitative data analysis procedures to merge the results of all the data collected from the different instruments.

Quantitative data obtained through test scores and check list questionnaire was analyzed using percentages, identifying and organizing the responses according to their similarities and finally the results were narrated.

On the other hand, the qualitative data which is obtained through semi-structured interview were described inductively that aims to drawing a general conclusion on the basis of specific evidence (Best & Khan, 2003). Hence the data which were collected through aforementioned tool was described using words to induce the relevant interpretation of the data.

3.7. Ethical Considerations

First, for the formality and legality of data collection, a letter of cooperation from Jimma University-Department of English Language and Literature was written to the concerned offices or schools. And then, the objective of the study was explained to the school administrative bodies and then to the participants of the study. In addition to this, the participants of the study were asked to give appropriate and genuine responses for the research questions based on their consent.

One of the data gathering tools used in this study and consumed student respondents time was a reading comprehension test that was administered twice. Thus, with the consent of school administrative bodies for they were responsible for adjusting suitable time (conducive temperature time and free noise disturbance time to make the respondents free of the tiredness and fatigue) in taking the test and with students consent the sample tests were administered.

In answering the questionnaire, the researcher promised that all the student respondents' responses were remained confidential and anonymous. Also, with the consent of school administrative bodies and English Language Department head, the researcher used the 2014/15 English exam-papers from exam banks for document analysis.

Finally, the researcher took the approval letter from Ano and Sire Preparatory Schools for the confirmation that the researcher collected data in an ethical way and finally attached the letter to the research paper in the Appendix section for approval by the major advisor and the co-advisor.

CHAPTER FOUR

4. FINDINGS AND DISCUSSION

The main purpose of this study was to assess the construct-irrelevant factors that affect students' performance in reading comprehension test in English classes. Moreover, attempt was made to address the following three research questions,

1. What are the major construct-irrelevant factors influencing students' performance in reading English comprehension test?
2. What are the adverse effects of the construct-irrelevant factors on students' performance while they are taking reading English comprehension test?
3. What are the strategies used by teachers or test-developers to overcome the construct-irrelevant factors affecting students' performance in reading English comprehension test?

Therefore, the findings and discussion section of this study deals with the analysis and interpretation of the data that were obtained through tests, questionnaires, document analysis and semi-structured interview in line with these objectives. This chapter is divided into three sections- firstly, the quantitative data obtained through tests, questionnaires, and a few related qualitative data and document analysis were analyzed jointly under successive similar topic to answer the first research question. Secondly, some qualitative data collected via interview, and quantitative data gathered through questionnaire were used to deal with the adverse effects of construct-irrelevant factors on students' performance in reading comprehension test. Similarly, the third part, with the quantitative data via questionnaire and qualitative data gathered through semi-structured interview were discussed and presented concerning with the strategies used to overcome the construct-irrelevant factors that affects students' performance on reading comprehension test.

4.1. Construct-irrelevant Factors Influencing Students' Reading Comprehension Test

To assess the factors affecting students' performance in reading comprehension test beyond the skill, ability, and knowledge that the test is intended to measure, the data which were obtained through four kinds of data gathering tools namely, test, questionnaire, document analysis, and semi-structured interview were analyzed jointly under their respective similar topic. The data collected through the questionnaire was developed based on the review literature and research questions. Additionally, teacher made or unstructured classroom tests were developed by the English teachers with the researcher to answer the language problems used in test questions by the English teachers. To end this, some related data found through semi-structured interview and document analysis (English exam papers) questions were analyzed qualitatively.

4.2. The Organization of Data Obtained from the Four Major Instruments

Ten questions based on the four essays (descriptive, expository, argumentative and narrative) were designed to collect data from language factors that influence test-takers and lower their performance in reading comprehension test. The tests were administered twice with the first session that contains unfamiliar languages (vocabularies) in the test questions and the second session contains the test languages (familiar vocabularies) to see the disparity of language factors on students' performance. All the test questions are short answer open-ended questions. To keep the reliability of these tests, the researcher used two scorers or raters from the two schools and analyzed each result quantitatively.

In addition, 19 questions from checklist questionnaire were divided into their respective construct-irrelevant factors under the following topics, namely, language factors, test accommodations factors, item/test format factors and the badly typed texts and test-questions factors and were analyzed quantitatively. And also the open-ended questions from the checklist questionnaire that contain the respondents' suggestions were organized thematically or qualitatively under these respective four headings.

Finally, six interview questions which were prepared for the teachers' respondents and five document analysis questions from the exam-bank were analyzed thematically under the major construct-irrelevant variables.

Table 4.2.1: Questions Analyzed under Different Construct-irrelevant Factors

Construct-irrelevant variables	Roll number of items in the checklist questionnaire	Total number of questions in the checklist questionnaire
Language related factors	10	1
Test accommodations related factors	7,8	2
Item/test format related factors	9	1
The badly typed and test-questions factors	1,2,3,4,5,6	6

Table 4.2.1 clearly presents the number of questions that were analyzed under different construct-irrelevant variables that affect students' performance in reading English comprehension test.

Table 4.1.2: Questions Analyzed under the Adverse Effects of Construct-irrelevant Factors

Construct-irrelevant variables	Items in the checklist questionnaire	Total number of Questions in the questionnaire
The usage of unfamiliar language	14	1
Inappropriate usage of test accommodations	12	1
The unfamiliarity of item/test format	11	1
The badly typed texts and test-questions	13,15	2

Table 4.2.2: Questions Analyzed as Strategies to Overcome Construct-irrelevant Factors

Construct-irrelevant variables	Number of items in the checklist questionnaire	Total number of questions
Unfamiliar language	18	1
Inappropriate usage of test accommodations	16	1
The unfamiliarity of item/test format	19	1
The badly typed texts and test-questions	17	1

On the other hand, students' suggestions from the questionnaire checklist like Q1, Q2, Q3, Q5, Q6, Q13, and Q15 were used for analyzing the adverse effects of badly written texts on students' reading English comprehension test. And questions like Q4, Q7, Q8 and Q12 were used for analyzing the adverse effects of inappropriate usage of test-accommodations in reading comprehension test. Similarly, Q9 and Q11 were used for analyzing the adverse effects of students' unfamiliarity with short-answer open-ended items in reading English comprehension test or item format. Finally, Q10 and Q14 were also used for analyzing the adverse effects of language (vocabulary) factors that influence the reading comprehension test. These all data were grouped under similar topics; and then the suggestions were thematically organized and described qualitatively

Therefore, in a similar manner, six questions were used for teacher respondents using semi-structured interview that dealt with the problems related to construct-irrelevant factors and its adverse effect on students' test performance and strategies used to alleviate the problems. And additionally, 5 open-ended questions were prepared and analyzed the reading comprehension test prepared in 2014/5 from the two schools which are available in the exam-bank (which services as document analysis). This was so as to triangulate the data obtained from the aforementioned data gathering instruments.

4.3. The Factors Related to Language

Ten open-ended reading comprehension questions were designed to collect data on how the unfamiliar languages or vocabularies on reading comprehension questions were affect students' performance in English classroom tests. Thus, the data obtained from students' test scores were discussed and explained successively as they appear in the two tables (Table 4.4 and Table 5.5) below. The discussions and explanations of the data involved comparing the results of the two test scores of students test prepared using unfamiliar languages (vocabularies) with the other test prepared with familiar languages (test languages). Since the item of the test is open-ended, its reliability is checked using Kappa (inter-rater reliability). In relation to this, Wang (2009) states that Kappa (K) is used for assessing inter-rater reliability and helps to determine consistency among raters in subjective tests. Therefore, the tests scored by the two raters were presented in the following ways for the analysis.

Table 4.3.1: Students' Test Scores by the Scorer "A"

Test scores with unfamiliar languages				Test scores with familiar languages			
Interval	Number of students(f)	Relative Frequency		Intervals	Number of Students(f)	Relative Frequency	
		Decimal	%			Decimal	%
10-22	8	0.18	18%	10-24	2	0.05	5%
23-34	5	0.11	11%	25-38	4	0.09	9%
35-46	14	0.32	32%	39-52	17	0.39	39%
47-58	11	0.25	25%	53-66	11	0.25	25%
59-70	5	0.11	11%	67-80	3	0.07	7%
71-82	0	0	0%	81-94	4	0.09	9%
83-94	1	0.02	2%	95-100	3	0.07	7%
Total	44	0.99~1	99~100%	Total	44	1	100%

The test scores from the table 4.3.1 shows the difference of students' test performance due to the language factors used in test questions. Thus, the percentages of students who scored low results in the test that contains unfamiliar language is high when compared to the percentages of students who scored low result in the test that contains the familiar language (test language). Also the percentages of students who scored high result in tests with the familiar language are higher than the percentages of students who scored high results with the test of unfamiliar language. This difference is occurred due to the difficult vocabularies, confusing, and abstract expressions used in test questions. (See Appendix C). In relation to this, Kopriva (2000) states that the inability to use test language (familiar language) by test-developers or teachers during English reading comprehension test when vocabulary is not tested contextually is a challenge for students' performance on tests that in turn affects students' test scores or results. In addition, in reading comprehension test, if the language of presentation is harder to understand than the passage itself, the reader is presented with an additional layer of difficulty and we cannot tell poor performance is due to the passage difficulty (Alderson, 2000).

Table 4.3.2: Students' Test Scores Rated by Scorer "B"

Test scores with unfamiliar vocabularies				Test scores with familiar vocabularies			
Intervals	Number of students(f)	Relative frequency		Intervals	Number of students(f)	Relative frequency	
		Decimal	%			Decimal	%
10-22	8	0.18	18%	15-27	3	0.06	6%
23-34	9	0.21	21%	28-39	2	0.05	5%
35-46	12	0.27	27%	40-51	14	0.32	32%
47-58	10	0.23	23%	52-63	13	0.30	30%
59-70	4	0.09	9%	64-75	5	0.11	11%
71-82	0	0	0%	76-87	4	0.09	9%
83-94	1	0.02	2%	88-99	3	0.06	6%
Total	44	1	100%	Total	44	0.99~1	99~100%

The test scores from table 4.3.2 shows the difference of students' test performance due to the language factors influencing students test scores. In addition to the data presented in Table 4.3.2, to see the reliability of test results in open-ended reading comprehension test, two raters (inter-rater reliability) was used. Thus, the percentages of students who scored low results in tests with unfamiliar language is higher than the percentages of students who scored low results in tests presented with familiar language. In addition, the percentages of students who scored high result in the test that contains familiar language are higher than the percentages of students who scored high results in tests that are provided with unfamiliar language. This is because the difficult vocabularies and vague expressions used in test questions affect students' language test performance. Therefore, the results indicated in Table 4.4 and this Table 4.5 is nearly the same.

Table 4.3.3: Descriptive Statistics for the Test Results of Students

	Rater "A" from Ano Preparatory School		Rater "B" from Sire Preparatory School	
	Test with unfamiliar languages	Test with familiar languages	Test with unfamiliar languages	Test with familiar languages
Statistics	First session	Second session	First session	Second session
N	44	44	44	44
Mean(X)	40.27	55.91	38.98	55.64
SD	16.640	20.008	16.925	19.404
Range	75	85	77	83
Sum	1772	2460	1715	2448

Key: N= total population of student respondents, SD= Standard Deviation

Table 4.3.3 shows that the descriptive statistics of students test result scored by the two raters. Thus, the mean score of students test result presented with the test of unfamiliar

languages for both raters is 40.27 and 38.98 respectively but the mean score of tests with familiar language is 55.91 and 55.64 respectively. On the other hand, the range of students test scores presented with difficult vocabularies is almost the same which is 75 and 77 and that of the tests presented with difficult vocabularies is 85 and 83.

The result of data from this descriptive statistics of the tests with familiar languages and unfamiliar languages indicates that the two tests are different in central tendency (as indicated by the means) which is 40.27, 38.98 and 55.91, 55.64 respectively and even more so in dispersion (as indicated by the standard deviations and ranges). In relation to this, Brown (1996) states that standard deviation is a means to estimation of the dispersion of scores and a good indicator of the dispersion of a set of test scores around the mean. It is therefore, it provides information on average distance that each score in a distribution is away from the mean. A larger standard deviation implies more variability in scores and a smaller standard deviation implies that the scores are close to each other or more similar (MOE, 2002). Thus, the result indicates that the test with unfamiliar language is small in standard deviation which means all the respondents test scores are near to each other. It means that the test is difficult for all respondents. On the other hand, the mean score of respondents test scores in the test of familiar language is relatively large. This indicates that there is a discrepancy of test scores among the test-takers. Therefore, the test questions presented with unfamiliar language (vocabularies) affects students' language test performance.

Table 4.3.4: The Symmetric Measures of Inter-Rater Reliability of the Two Scorers

Test with unfamiliar language scored by rater “A” * test with unfamiliar language scored by rater “B”				
	value	Assmp. Std Error ^a	Approx. T ^b	Approx. Sig.
Measures of agreement (Kappa)	.275	0.72	7.517	.000
Test with familiar language scored by rater “A” * test with familiar language scored by rater “B”				
Measures of agreement (Kappa)	.385	.077	9.232	.000

Key: * asterisk comparison, Value indicates the value of measures of agreement

Table 4.3.4 shows measures of agreement (Kappa) between two scorers in rating the short item questions taken from the reading comprehension test. Thus, items such as short answer questions require subjective scoring. In such cases, Kappa is used to measure reliability among the scorers which has a range from 0-1.00 with a larger value indicating better reliability. In relation to this, Wang (2009) states that Kappa is used to assessing inter-rater reliability and helps to determine consistency among raters in subjective tests. Therefore, the inter-rater reliability (Kappa) of the tests of unfamiliar language for both scorers is .275 which is interpreted as a fair agreement between the two scores. And also the inter-rater reliability (Kappa) of the tests of familiar language (vocabularies) for both raters is .385 which is interpreted as a fair agreement between the two raters. In relation to this, a more complete list of how Kappa might be interpreted according to Landis and Koch (1977) is < 0 poor agreement, 0.0-0.20 slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, and 0.61-0.80 substantial agreement, 0.81-1.00 almost perfect agreement. Therefore, the inter-rater reliability of the two test scores is interpreted as a fair agreement between the two scorers. This indicates that subjective scoring between the raters causes unreliability of test results. Therefore, from the researchers’ experience, the test-developers and scorers should take care for the bias resulted from the

quality of test-takers hand writing and neatness in answering the open-ended short answer questions and should stick to the ideas quality to the answer and appropriate weighting for the question points by using key answers.

Table 4.3.5: Correlation between the Test Scores with Unfamiliar Languages and Familiar Languages

	Tests	Correlation coefficient
Scorer “A”	Tests with unfamiliar languages and test with familiar languages	0.01
Scorer “B”	Tests with unfamiliar languages and test with familiar languages	0.01

Table 4.3.5 shows the correlation between the test scores of unfamiliar languages and familiar languages. Therefore, the correlation is significant at the 0.01 level. In relation to this, Marczyk, DeMatteo, and Festinger (2005) state that tests or measures of correlation will provide information on whether the correlation is statistically significant. Thus, one of the most widely used correlations is the Pearson product-moment correlation, often referred to as the Pearson r . which is used to examine associations between two variables that are measured on either ratio or interval scales. Thus, they also state correlation of .01 to .30 are considered small; correlation of .31 to .70 are considered moderate; correlation of .71 to .90 are considered large; and correlation of .91 to 1.00 are considered very large significance or strength of relationship (whether negative or positive). Therefore, the results indicated that the correlation of tests presented with unfamiliar languages to tests presented with familiar languages is 0.01 which lies between 0.01 to 0.03 is considered to be a small relationship. That is, there was a significance difference between test with unfamiliar language and test with familiar language on students’ test scores due to the difficult language factors and accessible language factors variation caused on student’s performance.

Table 4.3.6: Results of the Paired Sample T- test between the Two Scores on Unfamiliar and Familiar Languages

Paired Sample T- test									
		Paired Differences					T	Df	Sig. (2-tailed)
		Mean	S.D	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Test with unfamiliar language by rater A - test with familiar language by rater A.	-15.636	12.688	1.913	-19.494	-11.779	-8.175	43	.000
Pair 2	Test with unfamiliar language by rater B. - test with familiar language by rater B	-16.659	12.702	1.915	-20.521	-12.797	-8.700	43	.000

Looking at the mean score merely one cannot identify the significance difference between the test scores of students with unfamiliar language and familiar language tests. Therefore, from among the T-tests, the paired sampled t-test is used to compare two related means. It is also used to compare two population means when there are two samples in which observations in one sample can be paired with observations in the other samples or the before and after observations on the same subjects (Park, 2009).

Thus, from table 4.3.6, the mean differences of the two tests of the same respondents are negative because the result of test with familiar languages is subtracted from the results of test with unfamiliar languages. So the mean difference shows that the test scores tended to increase more with the familiar language tests.

4.3.7. The Factors Related to Questions Containing Unfamiliar Languages

Question number 10 from the checklist questionnaire was designed to collect data on how the language (unfamiliar words and expressions) in the test question affects students' test performance in reading comprehension test. The presentation and discussion of data involved comparing the percentages of students' responses from the table and thematically organize their suggestion from the questionnaire checklist provided for them.

Table 4.3.7.1: Difficult Language Factors in Reading Comprehension Test Questions

Number	Questionnaire	Yes	100%	No	100%
Q10	Are the test questions sometimes provided with difficult vocabularies?	37	84%	7	16%

As it can be seen from table 4.3.7.1, 37 (84%) of the student respondents reported that the test questions prepared by test-developers sometimes provided with difficult vocabularies. In relation to this, Alderson (2000) stated that in a reading comprehension test, if the language of presentation is harder to understand than the passage itself, the test taker is presented with an additional layer of difficulty in doing the test. Thus, the usages of unfamiliar vocabularies instead of test-languages affect students' test performance in reading comprehension test. On the other hand, 7(16%) of student respondents reported that the test questions were not presented with difficult vocabularies because as they suggested, it is the nature of the language exam presented in this way. But the majority of student respondents suggested that it is the difficult vocabularies provided in test-questions affect their test performance.

In order to crosscheck the findings, the researcher conducted interview with 5 English teachers' participants. The participants were asked to forward their suggestions with respect to the language factors that affect students' performance on reading

comprehension test. Thus, 3 of the five teachers reported that the careless construction and development of tests by test developers and the inappropriate usage of words (languages) that are not commonly used to ask questions from the reading passage can affect test-takers performance and the text (reading passages) which does not initiate them or remote from them socio-culturally are full of messages and strange words beyond their abilities can affects their performance.

In addition, the other two teachers suggested that the usage of academic languages used for testing the other subjects rather than language testing particularly reading English comprehension test can affect students' language test performance. The words such as "discuss and explain" can be ambiguous. Therefore, if the test-developers use "discuss and explain" then they must give specific instructions as to what points should be discussed and explained otherwise it affects students' performance on tests.

Finally, the analysis of reading passages and test questions from the exam-bank indicated that there are some questions which are presented with strange vocabularies and difficult expressions which/that are beyond the ability of the students which in turn affects students test performance. For instance, the word laity in Appendix "M" is a strange vocabulary for the test-takers and its meaning is not inferred from context. The difficult expression for example, "by the same token", and the word such as "invaluable or priceless" is presented in test questions but not in context.

4.4. Test Accommodations Factors Influencing Students' Performance

The following two questions (Q7 and Q8) were designed to collect data on the test-accommodations that influence students' performance on reading comprehension test.

Table 4.3: The Test-Accommodation Related Factors in English Language Test

Number	Question	Yes	100%	No	100%
Q7	Is the correction orientation on test-questions and the reading passages given by:				
	a. The invigilator during the test?	16	36.4%	10	22.7%
	b. The subject teacher during the test?	3	6.8%	7	15.9%
	c. Sometimes by both invigilator and subject-teacher?	3	6.8%	5	11.4%
Q8	Is disruptive behavior sometimes revealed from the invigilator during the test?	37	84%	7	16%

Table 4.4.1, depicts that 16 (36.4%) of student respondents reported that correction orientation of the test questions and the reading passages are given by the invigilator. As they suggested the correction orientation given by the invigilator is not as much as accurate like that of the subject-teacher so that it affects their test performance.

Seven (15.9%) of student respondents suggested that correction orientation of test-questions are not given by the subject teacher during the test. This is because as they suggested the subject teacher advised the invigilator to give correction orientation of test-questions and the passages during the test on the behalf of the subject teacher so that the invigilator is not efficient to correct the error occurred in the test. Therefore, it is one of the test-accommodation factors that affect student's language test performance. On the other hand, 5(11.4%) of student respondents argued that correction orientation of test-questions and the reading passages are sometimes not given by both invigilator and subject-teacher. Thus, as they suggested the error is much and it is difficult to correct within a short period of time. This in turn affects students' language test performance.

In other cases, most of the students in check list questionnaire suggested that when the correction orientation was given by invigilator or the subject teacher, it distorts test time because the time for correction orientation is not considered and not allocated in the

normal test time. This is therefore; sometimes the corrections are given at the end by neglecting the appropriate time allotment for the test. Thus, helping students with correction orientation and allocating appropriate time for the test and appropriate time for correction orientation is an appropriate test accommodation.

In addition, 37(84%) of student respondents reported that the disruptive behavior sometimes revealed from the invigilator. This is because as they suggested the test-takers are asking questions again and again for the problem encountered them on tests and this leads the conflict between invigilators and test-takers which in turn affects test-takers' time in doing the test. In addition, they also suggested that the disruptive behavior from some students in using mobile phones for cheating is another source of conflict between the invigilator and students. On the other hand, 7 (16%) of student respondents reported that the disruptive behavior is not revealed from the invigilators and they argued that it is the behavior revealed from invigilator to control exam cheating.

Seeking further clarification the researcher used interview to cross-check the finding with the teacher interviewee to suggest their ideas about the kind of situations and test-taking activities that affect students' test performance. Thus, they argued that the disruptive behavior sometimes revealed from some invigilators influence students' performance. With respect to this question, three of the teachers' respondents suggested that the intimidation from some invigilators due to some test-takers usage of mobile phones, the inability of the invigilator to stabilize students striving for asking questions again and again in need of support creates conflict among the test-takers and invigilators. In addition, the neglecting of time assigned for correction orientation which consumes test-takers time to do the test, absence of the provision of separate answer sheet with reserves are some of the situations and test-taking activities that cause conflict between the invigilator and test-takers affect their performance in relation to test accommodations.

Finally, the document analysis of reading passages and test-questions from the exam-bank indicated that there are some questions and alternative choices provided on different pages of question papers that create confusion and consumes students' time to do the test.

In addition, the data obtained from the document analysis indicated that the alternative choices presented in the form of “all of the above”, “none of the above”, A & B and also there is the usage of alternatives by the test-developers from A-E on the classroom reading English tests. According to JUCDE (2014, p.104), this form of presentation is against the rule of Brown’s law that explains “If one distracter is ridiculous, that distracter is not helping to test students. Instead, those students who are guessing will be able to dismiss that distracter and improve their chances of answering the item correctly without really knowing the correct answer.” Thus, most test-developers want their tests to be relatively efficient by adding extra-information that is not related to the construct being tested. This will just take extra-time for students to read and will add nothing to the test.(See Appendix M).

4.5. The Item/test Format Factors on Students’ Language Test Performance

Question number 9 was designed to collect data on how the students’ unfamiliarity with short-answer open-ended questions from the reading comprehension test affects their test scores.

Table 4.5.1: The Factors Related to Students’ Unfamiliarity with Item/test Format

Number	Questionnaire	Yes	100%	No	100%
Q9	Have you experienced open-ended reading comprehension test in English classroom tests?	4	9%	40	91%

As it can be seen from table 4.5.1, 40 (91%) of student respondents reported that they are not experienced the short-answer item in answering the reading comprehension test. In a checklist questionnaire, they also suggested that they experienced multiple-choice items on reading comprehension tests and this in turn affects their performance because of the experience they sometimes encountered in classroom reading comprehension exercises in giving short-answers orally not by writing. And also the influence of active learners who

dominate in answering the oral questions in classroom exercises of reading comprehension test hampers the chances of the majority of learners from participation and experiencing the questions. In relation to this, Fairbairn (2007) states students who are unfamiliar with the format of the test itself or the test questions may likely have difficulty in demonstrating what they know and are able to do on tests.

The teacher respondents were interviewed if they have noticed any particular item on a test rather than multiple-choice items. Regarding this question, all of the teacher respondents stated that the most common item that the test-takers are familiar with in reading comprehension test is multiple-choice items. They also suggested that the test-developers or teachers construct test questions from the reading passage using multiple-choice items because it is easy to score. Thus, students are unfamiliar with short-answer open-ended reading English comprehension classroom tests. Therefore, this is one of the factors that affect students test performance in reading English comprehension test.

Finally, the data from the passage and comprehension questions from the exam-bank using the document analysis indicated that all the questions are prepared with the multiple-choice items not with short-answer items. Therefore, this indicates that the questions taken from the passage is always prepared with multiple choice items in the two schools.

4.6. The Factors Related to Badly Typed Passages and Test Questions

The following questions, Q1, Q2, Q3, Q4, Q5 and Q6 were designed to gather data on how the badly typed texts and test questions affect students' test performance in English classroom tests.

Table 4.6.1: The Badly Typed Texts and Test Questions Factors

Number	Questions	Yes	100%	No	100%
Q1	Are the texts (passages) and comprehension questions always well-spelt and clear in terms of:				
	a. Spelling	5	15.4%	19	42.3%
	b. Punctuations	4	9.1%	8	18.2%
	c. Capitalization	2	2%	6	13.6%
Q2	Is the exam paper neat and attractive?	6	14%	38	86%
Q3	Are the initials of each and every paragraph of the reading passages assigned with numbers?	3	7%	41	93%
Q4	Is vocabulary in context asked for its meaning always written in bold letters?	13	29.5%	31	70.5%
Q5	In multiple-choice items of test-questions, are the questions and alternative choices sometimes provided on separate pages?	32	73%	12	27%
Q6	Are the alternatives provided in the form of				
	1. A. _ B. _ C. _ D. _ or (Horizontally)	20	45.5%	7	15.9%
	2. A. __ C. __ or (in column) B. __ D. __	4	9.1%	6	13.6%
	3. A. _____ B. _____ or (Vertically) C. _____ D. _____	4	9.1%	3	6.8%

As shown in table 4.6.1, question number 1, 19(43.2%) of student respondents reported that the reading passages and questions taken from the passages are not written with correct spelling and 8(18.2%) of student respondents reported that there is a problem of

punctuation in the given reading comprehension test and finally 6(13.6%) of student respondents reported that the passage and test-questions are presented with wrong capitalizations. In relation to this, Hughes (1989) states if the texts and comprehension questions are not well spelt and clear, it creates variations on students' performance by causing unwanted tasks which result in lowering test-takers test scores. Therefore, the passages and test questions presented with an error in spelling is the main factor that cause students test performance in doing the test followed by punctuation and capitalization respectively.

With reference to question number 2, 38(86%) of student respondents reported that the passage and comprehension questions are not as much as neat and attractive. In relation to this, the majority of students suggested that the passages and comprehension questions are not as much as neat and attractive due to the invisible and disorganized written passage and questions and the usage of inappropriate inks in writing and printing the exam papers. They were also asked to forward their suggestions and the vast majority of them suggested that there is the usage of too much text (reading passage) and test questions in too a small space during the classroom tests. In relation to this, Hughes (1989) states classroom tests are badly typed, and having too much text in too a small space and poorly reproduced. As a result, test-takers are faced with additional tasks which are not once meant to measure their language ability.

In Table 4.6.1, question number 3, 41(93%) of student respondents reported that in English classroom tests each paragraph of the text is not assigned or presented with numbers. The vast majority of student respondents suggested that the initials of each paragraphs of a given texts are not presented with numbers and this in turn affects students' time in finding the number of paragraphs. In this case, when they search for information from the given paragraphs to answer the question, the test-takers are required additional layer of difficulty. Therefore, the paragraph not assigned with number is a challenge for test-takers in identifying the paragraph from which the information (answers) for the given question is found.

As shown in table 4.6.1, question number 4, 31(70.5%) of student respondents reported that the vocabulary in context asked for its meaning from the passage is sometimes not written in bold letters. The majority of them also suggested that the vocabularies or phrases asked from the text (passages) are written with italics. But, 13(29.5%) of student respondents reported that the vocabularies asked for its meaning from the text are written with bold letters. They also suggested that the words or phrases written in italics were considered to be bold, so it is said that the words or phrases are written in bold letters. Thus, the word or phrase asked for their meanings are not easily visible to the test-takers so that their time is distracted in searching such words and phrases from the passage in doing the test.

Table 4.6.1, question number 5, depicts that 32(73%) of student respondents reported that the questions and its respective alternative choices in multiple-choice items are sometimes provided on separate pages. That is, the questions and the choices are presented on separate pages. Thus, it creates confusion and hesitation on test-takers in finding it from another page.

As shown in Table 4.6.1, question number 6, 20(45.5%) of student respondents reported that the alternatives in multiple-choices are written horizontally rather than vertically or in column. On the other hand, 6(13.6%) of student respondents reported that the alternative choices in the reading comprehension test are not presented in the form of columns. And finally, 4(9.1%) of student respondents reported that the alternative choices in reading comprehension test is not always presented in vertical ways. Therefore, it is acceptable to write the alternative choices in the form of columns for the shorter ones and verticals for the longer ones in order to be seen by the test-takers to enhance their performance.

The student respondents were also asked to suggest their opinion and the majority of them suggested that when alternative choices are not presented in the form of column and verticals, the questions are not easily visible and clear to the test-takers. In relation to this, Hacuta (2009) states that institutional tests are badly written and have too much text in

too a small space and are poorly reproduced with respect to fonts, font sizes, and line breaks in reading passages and test questions.

In order to cross-check the findings, the researcher conducted an interview with teacher respondents and all of them suggested that the usage of too much text in too a small a space due to the sanction imposed on test materials by the school administrative bodies in the way they managed the budget can affect students' performance on reading comprehension test.

Finally, the data from the exam-bank about the passages and comprehension questions using the document analysis indicated that there is the usage of too much text into a small a space so that there is an error in spelling and punctuations. The questions and its corresponding alternative choices are written in the same line horizontally which is difficult for the test-takers to identify it easily and the alternative choices are not written below the questions in an indented form from the test questions. In addition, the initials of each paragraph of the texts (reading passages) are not assigned with numbers and sometimes the same paragraph is split into two without conveying a clear and full message. In another case, sometimes, different paragraphs are connected with another paragraph which is difficult to identify by the students. Therefore, for the alternatives that contains shorter choices, columns and for the alternative that contains longer choices presenting the alternatives in vertical ways is advisable.

4.7. The Adverse Effects of Unfamiliar Languages on Students' Test Performance

The following one question (Q14) of the questionnaire presented in table 4.7.1 was generated for data gathering tool concerning the adverse effects of unfamiliar vocabularies on students' language test performance on reading comprehension test.

Table 4.7.1: The Adverse Effects of Unfamiliar Languages on Students' Performance

No.	Questionnaire	Yes	100%	No	100%
Q14	Do you feel panic when you see the difficult vocabularies used in questions during reading comprehension test?	38	86.4%	6	13.6%

Table 4.7.1, depicts that 38(86.4%) of the student respondents feel panic when they read the difficult vocabularies while they are doing the reading comprehension in English classroom test and 6(13.6%) student respondents never feel panic when they survey (look at) difficult vocabularies used in reading comprehension test. In a checklist questionnaire students are asked to forward their suggestion and the majority of them suggested that they feel panic when they see the difficult vocabularies used in test questions during the reading comprehension test. Therefore, difficult vocabularies presented in test questions cause students feel panic which in turn affects test-takers performance.

Regarding the adverse effects of unfamiliar languages on student's performance in reading comprehension test, all of the teacher respondents suggested that confusion and the hesitation on the side of test-takers are the adverse effects that are caused by strange vocabularies which in turn cause students' low performance on tests.

4.8. The Adverse Effects of Test Accommodation in Reading Comprehension Test

Question number 12 was designed to collect data about the adverse effects of test-accommodation on students' language test performance in English classroom tests. The data obtained from Q12 was discussed and presented as follow.

Table 4.8.1: The Adverse Effects of Test-Accommodation on Test-takers

Number	Questionnaire	Yes	100%	No	100%
Q12	Do you feel angry so that you hate the subject teacher due to the assignment of too little time allocation for the test that needs too much time?	35	79.5%	9	20.5%

The response to question 12, in table 4.8.1, indicated that 35 (79.5%) student respondents feel angry so that they hate subject teacher. As they suggested this is because they feel tense due to the time constraint they encounter on tests. As a result, they feel angry and hate the subject and subject teacher. On the other hand, 9 (20.5%) of the respondents never feel angry and hate the subject teacher. As they suggested they consider the problem as easy and will be corrected easily. In a checklist questionnaire students are asked to forward their suggestions and they suggested that they feel angry, hate the subject and the subject teacher because of the assumption that the causes for the problems are the subject teacher.

Regarding the adverse effects of test-accommodations all of the teacher respondents expressed that due to the time constraint students are carelessly answering the questions to cope with the time to do the test and therefore, they blame the subject teacher. In short, carelessness to do the test and blaming the subject teacher is the adverse effects of test-accommodation.

4.9. The Adverse Effects of Unfamiliar Item/Test Format

The following table, table 4.9.1 depicts the influence of students' unfamiliarity with the item/test format in answering short-answer open-ended questions from reading comprehension test.

Table 4.9.1: Students' Lack of Experience in Answering Short Answer Questions

Number	Question	Yes	100%	No	100%
Q11	Do you feel worry for your test scores due to your unfamiliarity with the test format?	35	80%	9	20%

Table 4.9.1, depicts that 35 (80%) of student respondents feel worry for their test scores. They also suggested that they feel worry and think because of the fear of very low test scores, fear of their families, and fear of the sentiments of their classmates and their subject teachers and this in turn distracts their attention from doing the test. But, 9(20%) of student respondents reported that they never feel worry for their test scores. This is because as they suggested it is the problem of unexpected test item we encountered on tests beyond the experience we accustomed previously. Therefore, test-takers unfamiliarity with the open-ended questions cause students feel worry for their test scores.

In a checklist questionnaire the students were asked to forward their suggestion and some of them suggested that they worry for their test scores because they suspect fear of failure in the test.

On similar issues, the teacher respondents expressed that students loss of experience in answering short-answer questions has an adverse effect on students' performance in that it is not conducive for test-takers to guess the answer like that of multiple-choice items. Therefore, they think about their test score by leaving behind to do the test questions and ask the invigilator how to do the questions again and again.

4.10. The Adverse Effects of the Badly Typed Texts/Test Questions

The following questions, questions 13 and 15 were designed to gather data on how the badly typed texts and test questions affect students' language test performance in English classroom tests.

Table 4.10.1: The Adverse Effects of the Badly Typed Texts/Test Questions

No.	Questions	Yes	100%	No	100%
Q13	Do you feel panic when you see the bad typing format used in texts (passages) and test questions?	39	88.6%	5	11.4%
Q15	Do you get test anxious (fear of test) during the reading comprehension test due to the badly written texts and test questions?	41	93.2%	3	6.8%

Table 4.10.1 depicts that 39 (88.6%) of student respondents' stated that they feel panic when they survey (look at) the bad typing format of the reading comprehension test. On the other hand, 5(11.4%) of student respondents reported that they never feel panic by surveying the bad typing format of the test. Therefore, the badly typed texts or passages cause students feel panic. In question number 15, 41(93.2%) student respondents feel test anxious (fear of test) in taking the reading comprehension test. In relation to this, Hughes (1989) states the badly typed and too much text or test questions in too small space causes anxiety and affects student performance on language tests.

In a checklist questionnaire the students were asked to forward their suggestions what they feel when they see the bad typing format of the text and test questions while they were doing the reading comprehension test. Thus, the majority them suggested that they feel panic when they review the bad typing format which is difficult to read and answer the reading comprehension questions.

In table 4.10.1, for question number 15, the students were asked to forward their suggestions what they feel when they think about the test. Thus, they feel test anxious in reading comprehension test because they think about the too long and badly written texts lowers their test performance. In relation to this, Putwain (2008) states that test-anxiety occurs in an evaluative situations or contexts in which an individual’s fear performing poorly on tests because a poor score might hinder their academic goals or lead to ostracism from peers.

4.11. The Strategies to Overcome the Language Factors in Reading Comprehension Test.

Question number 18 was designed to collect data on how to overcome the language factors used in reading comprehension test by the test developers.

Table 4.11.1: Strategies used by Test-developers to overcome the Language Problems

Number	Questionnaire	Yes	100%	No	100%
Q18	Usage of plain language in test-questions by the test developers/teacher	41	93.2%	3	6.8%

As it can be seen from table 4.11.1, 41(93.2%) of student respondents suggested that the usage of plain language used in test questions by the test developers/teachers in developing classroom tests are used as the strategies to overcome the language factors affecting students language test performance in reading comprehension test. In relation to this, according to Kopriva (2000), the absence of “plain language” in test development by the test-developers lowers the students’ test performance. But, 3(6.8%) of student respondents suggested that the usage of plain language is not used as strategies to overcome the language problems.

In line with the above strategies obtained through test-takers questionnaire, the researcher used to crosscheck the finding using test-developers interview. Thus, one teacher respondent reported that the test-developers should use the most common words or phrases used to ask questions from the reading passage which serve as a basis for constructing classroom reading comprehension test. This includes the expressions related to skimming for the main idea or topic, scanning for details, guessing word meaning from context, and inference. In relation to this, Brown (2004,) states that the basis for constructing classroom comprehension test is based on skimming for the main idea and topics, inference (implied details), guessing word meaning from context, grammatical features for example, it, they and etc refers to, details (scanning for specifically stated details), and excluding facts not written (for example, which one of the following is not mentioned....) and using discourse markers with simple languages.

On the other hand, the other teacher respondent stated that the most common expressions and questions that are used to ask questions from the reading comprehension test probably include:

1. What do you think the main ideas of this passage?
2. What would you infer from the passage?
3. Which one of the following is not mentioned in....
4. The pronouns “It”, “they” etc refers to
5. According to the passage,
6. The expression “ ___ ” in line ___ refers to

The other two teachers suggested that skimming, scanning, inference and vocabulary from the context are the basis for asking questions and used as a basis for constructing classroom based reading comprehension test when care is given to the appropriate usage of plain language (simple language) and confusing words and expressions should be omitted. And they also suggested that the test-developers should use the common phrases

and words that are used to ask questions from the reading passage such as refers to; mean; guess, what do you infer ...?; what is the title ...?; what do you understand ...?; what this word mean...?; what is the main idea/central point or thesis...? to ask questions from the reading passage. In addition, the test-developer should prepare the test blue print (or test specification) which identifies the objectives and the skills which are to be tested and the relative weight on the test given to each because it provides clear instructions and a crucial first step in the test development process. Moreover, the short training on how to use and construct simple language in constructing test should be given to the subject teachers by the nearby university. Therefore, if the test-developers or teachers are equipped with such skills, the problem of using inappropriate language on tests can be alleviated.

4.12. The Strategies to Overcome the Problems of Test-Accommodations Factors

Question number 16 was designed to gather data about the strategies used by teacher or classroom test developers that encounter students' on reading comprehension test.

Table 4.12: The Strategies to Overcome the Inappropriate Usage of Test Accommodations

Number	Questionnaire	Yes	100%	No	100%
Q16	Usage of careful correction of the given passage and test-questions by subject teachers	42	95.5%	2	4.5%

As it is presented in Table 4.12, 42(95.5%) of student respondents reported that the usage of careful correction of the language and test-questions by subject teachers before the test is administered to the test-takers can alleviate the inappropriate usage of test accommodations. The vast majority of them also suggested that the careful correction orientation by the subject teacher prior to the test administration help them to use their time properly. In relation to this, Fairbairn (2007) states common accommodations such

as appropriate time allocation in a response mode on tests are appropriate for second language learners. Therefore, correction orientation is vital in saving student's time in doing the test.

On similar issues, the teacher respondents expressed that the time that the students require or need to answer each and every question should be taken into consideration while allocating time. Thus, the subject-teacher set aside time for correction orientation so that the time for correction by the subject-teacher and the time assigned for doing the test by the test-takers are clearly demarcated. Therefore, the correction given by the subject teacher within appropriate time is advisable.

4.13. Strategies for Coping with the Problems Related to Item/Test Format

Question number 19 was designed to gather data on how to overcome the problems of students' unfamiliarity in answering short answer open-ended questions from the reading comprehension test in English classroom tests.

Table 4.13.1: The Strategies to Overcome Item/Test Format Factors by Test-Developers

Number	Question	Yes	100%	No	100%
Q19	Familiarize students' in answering open-ended questions from reading comprehension test in English classroom test.	41	93.1%	3	6.9%

It is clearly indicated in Table 4.13.1, question 19, 41(93.1%) of student respondents suggested that it is possible to alleviate the problems of students' unfamiliarity with the item format by familiarizing students in answering short answer questions from reading comprehension test in English classroom tests. In relation to this, Popham (2005) states in classroom based tests, students are more likely to be able to best demonstrate what they know and can do if the test-takers are familiar with them. On the other hand, 3(6.9%) of

student respondents argued that it is not advisable to familiarize the test-takers to item/test question because it is the matter of students effort to answer the short answer open-ended question items from the reading comprehension test. The majority of the student respondents suggested that it is the responsibility of test-developers or teachers to familiarize students with the short answer questions so that students are familiarized and do the short answer questions with less difficulty. Therefore, the subject-teacher should familiarize test-takers to the short-answer items to alleviate the problems of students' unfamiliarity with the strange items of open-ended questions that they encounter on tests.

4.14. Strategies for Coping with the Problems of Badly Typed Texts and Test Questions

Question number 17 was designed to gather data on how to overcome the badly written texts and test questions that affects students' performance on reading comprehension test in English classroom tests.

Table 4.14.1: Strategies Used to Overcome the Badly Written Texts and Test Questions

Number	Question	Yes	100%	No	100%
Q17	Usage of experienced or computer literate in writing the reading comprehension test	37	84%	7	16%

As shown in table 4.14.1, question 17, 37(84%) of student respondents reported that using experienced type writer in reading comprehension test can alleviate the challenges of the badly written texts in classroom tests. On the other hand, 7(16%) of student respondents reported that using experienced type writer in writing the reading comprehension test cannot overcome the problem of badly typed written texts. They also suggested that it is the effort of the test-developers to guide the type writer in the provision of well written texts for the test-takers. Therefore, the usage of experienced or computer literate secretary in writing the reading comprehension test can minimize the badly written passages and test-questions.

With regard to badly written texts in the interview questions the two teacher respondents reported that test-developers should check the vocabulary in the context should be typed in underlined, or putting them in boldfaced type. On the other hand, the other two teacher respondents reported that the subject teacher is responsible to correct or guide and take proactive actions in correcting all the important things in the written reading comprehension test. Finally, one teacher respondent reported that the test-developer should assign each paragraph of a passage with numbers to be easily understood and identified by the test-takers and also the test should be edited and checked so as to free from all writing error beforehand.

CHAPTER FIVE

5. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

This chapter deals with the overall summary of the study, the conclusions of the findings and recommendations made based on the finding.

5.1. Summary

This study was conducted with the aim of assessing the construct-irrelevant factors that affect students' performance in reading English comprehension classroom test at Ano and Sire Preparatory schools of Grade 11 students. So as to deal with this objective, the following research questions were raised as follow:

1. What are the major construct-irrelevant factors affecting students' reading English comprehension test?
2. What are the adverse effects of the construct-irrelevant factors affecting students' performance while they are taking reading English comprehension test?
3. What are the strategies used by teachers or test-developers to overcome the major construct-irrelevant factors affecting students' performance in reading English comprehension test?

In answering such research questions, the mixed research design which encompassed both qualitative and quantitative research methods was employed. The questionnaires, tests, semi-structured interview, and document analysis (reading comprehension test were from the exam bank) were used to obtain data for the study. The quantitative data which collected via tests were analyzed quantitatively with the descriptive statistics to compute the mean, standard deviation, correlation coefficient, and inter-rater reliability (Kappa) and the sampled paired t-test using SPSS version 20. In addition, the quantitative data obtained through questionnaire were analyzed quantitatively. On the other hand, the

qualitative data obtained through interview and document analysis were analyzed inductively. The participants of the study were 44 students of Grade 11 and 5 English teachers of Ano and Sire Preparatory schools who were selected by purposive sampling. Having collected the data through aforementioned tools, the researcher discussed and presented the findings of the study.

The findings indicated that the careless development and construction of tests by the test-developers, the usage of too much text in too a small a space, the inappropriate usage of test accommodations such as time constraints and invigilators disruptive behavior, and language matters such as the inability to differentiate the test languages from the everyday languages and instructional or academic languages from the side of test-developers in constructing the test affects test-takers performance. In addition, the badly written texts (passages) and test-questions, students' unfamiliarity with the test-items and questions were found to be the major construct-irrelevant factors that influence students' performance in reading comprehension test.

With respect to the adverse effects of construct-irrelevant factors on students' performance, the findings also indicated that fear of failure, fear of test (test-anxiety), low test score, hating the subject matter, disliking subject-teachers, getting tense and panic, loss of hope in doing the short answer item in reading comprehension test and confusion were found to be the impacts of the construct-irrelevant factors that affect students' test-scores.

Finally, awareness creation for test-developers on how to construct the reading comprehension test and the factors considered in deciding test construction such as what is to be measured using the test specifications (blue print) by English department and the nearby universities, familiarize test-takers with the test-items, the usage of experienced type writer, correction and editing of the text and test question prior to test administration by the test developers and sufficient allocation and provision of test materials by the school administrative bodies were stated as the strategies that enable the test-developers or teachers to overcome the construct-irrelevant factors in reading comprehension test.

5.2. Conclusions

The following conclusions were drawn from the finding of the study.

1. The results of this study indicated that difficult vocabularies, abstract, vague expressions and the inability to use test languages by test-developers were found to be the major cause of construct-irrelevant factors that influence students' performance in reading comprehension test. The findings of the study clearly indicated that the usage of inaccessible words and expressions in test questions by the test-developers was found to be major factors that influence students' performance in reading comprehension test.

2. The ineffective usage of environment in which the assessment is administered, time allotment, additional materials such as separate answer sheet with reserves, invigilators inappropriate support of students during the test when the test-takers encounter problems were also a powerful source of construct-irrelevant factors that influence students' test performance. With respect to this, the vast majority of respondents suggested the point that variations in administrations such as time allocation, invigilator's disruptive behavior and sometimes his/her carelessness and lack of additional test materials affect students' performance in reading comprehension test.

3. Students' unfamiliarity with the open-ended reading comprehension test was the major construct-irrelevant factors that affect students' performance. This was happened due to the students' more familiarity to the multiple choice items in the classroom tests which provides a chance of guessing.

4. The badly typed texts and test questions such as the usage of too much text in too a small a space, paragraphs not assigned with numbers, the tests that are not well laid out and perfectly legible, vocabularies in context not written in bold letters, inappropriate written spelling and punctuations were also the major construct-irrelevant factors that affect students' performance. Therefore, the badly written texts and test questions are the major factors that affect student's performance in reading comprehension test.

5. Test-anxiety which result in getting panic, tense and worry for the test-score were found to be the major adverse effects of construct-irrelevant factors that affect students' performance in reading comprehension test. This result was occurred due to the badly typed texts and test-questions in reading comprehension test.

6. Hating the subject, subject teacher and loss of interest on the side of students to do the open-ended reading comprehension tests were also the adverse effect of construct-irrelevant factors which affect students' performance in reading comprehension test.

5.3. Recommendations

The following recommendations were made based on the findings of the study in order to alleviate the construct-irrelevant barriers that influence students' performance in reading English comprehension test.

1. Teachers or test-developers should identify test-languages from other inaccessible languages to ask questions from the reading comprehension test. That is teachers should use the commonly used words, phrases, and expressions in reading English classroom test. In addition to this, the provision of short training for language teachers how to construct a reading comprehension test by the English teachers from the nearby university is advisable.
2. Teachers should consider and facilitate a precondition for the appropriate usage of test-accommodations such as test-administration environment, time allotment, extra (additional) and separate answer sheet as reserves. The disruptive behaviors of the invigilator should be addressed with proper preventive measures in collaboration with English department and school administrative bodies so that the invigilator facilitates the testing conditions during the English reading comprehension test. Also during examination, invigilators should know what to do if regulations are broken, or there are other events, foreseen or unforeseen such

as when test-takers are found cheating, when there is a powerful failure or when something else happens that may cause bias or unfairness or force the session to stop. In the case of cheating, invigilators should be aware of the possible dangers from devices such as mobile phones with cameras. Therefore, before the beginning of the exam, clear instructions should be given to test takers how to behave during the examination by the invigilators. This may include information about unauthorized materials, the use of mobile phones, leaving the room during the exam and the start and end times.

3. The subject-teachers should orient students how to negotiate the open-ended short answer questions and familiarize them by practicing the tests in reading comprehension test or exposing students to various test formats.
4. The teachers or test-developers should use neat and clear question papers, use experienced typewriters, careful editing of the texts and test-questions, fonts, font sizes, line breaks, spelling, and punctuation.
5. The test-developers should provide continuous or frequent assessment tests with open-ended short answer reading comprehension tests which are well laid out and perfectly legible so that they feel confident rather than getting test-anxious.
6. The test-developers or teachers provide correctly written texts (reading passages) that initiate the test-takers and consult them its advantage in their future academic testing so that intimacy of experiencing the open-ended questions will develop gradually.
7. The study paves the way for the other researchers to conduct a study on construct-irrelevant factors in reading English comprehension tests and other tests.

References

- Abedi, J. (2002). *Assessment and Accommodations of English Language Learners. Issues Concerns, and Recommendations. Journal of School Improvement, 3*, Retrieved September 2, 2014 from [http://www.nacacasi.org/jsi/2002v3il/assessment adp](http://www.nacacasi.org/jsi/2002v3il/assessment%20adp).
- Abedi, J. (2006). Language issues in item development. In S.M. Downing & T.M. Haladyna (Eds.). *Handbook of test development*: Mahwah, NJ; Lawrence Erlbaum Associate.
- Abedi, Courtney, Mirocha, Leon & Goldberg. (2005). Language accommodations or English language learners in large-scale assessments: *Bilingual Dictionaries and Linguistic Modifications (CSE Technical Report 666)*. Los Angeles, CA: Center for the study of Evaluation/National Center for Research on Evaluation, Standards, and Student Testing. Retrieved from <http://WWW.cresst.org> on October 24, 2014.
- Abedi, J., Lord, Hofstetter, C., & Baker, E. (2000). The impact of communication strategies on English language learners' test performance. *Educational Measurement Issues and Practice, 19*(3).
- Abedi, J., Lord, C., & Palmer, J.R. (1997). Final report of language background as a variable in NAEP mathematics performance Angeles (CSE Technical Report 429). Los, CA: Center for the study of Evaluation/National Center for Research on Evaluation, Standards, and student testing. Retrieved from <http://WWW.cresst.org> on October 28, 2014.
- Alderson (2000). *Language assessment series*. Cambridge: Cambridge University Press.
- Alderson, J. C., Clapham, C. & Wall, D. (1995). *Language test construction and evaluation*. New York: Longman.
- Alderson, J. C., Clapham, C. & Wall, D. (2000). *Language test construction and evaluation*. Beijing: Foreign Language Teaching and Research Press.
- Alemayehu, N. (2008). A study of the practice of integrating language skills in the teaching of English (Unpublished MA Thesis). Addis Ababa University.
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.

- Bachman, L.F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language testing*, 17.
- Bachman, L.F. & Palmer, A.S. (1996). *Designing and developing useful language tests*. New York: Oxford University Press.
- Bailey, A. L. (2007). *The language demands school: Putting academic English to the tests*. New Haven, CT: Yale University Press.
- Barrett, T. (1976). Taxonomy of reading comprehension. In R. Smith and T. Barrett (Eds.). *Teaching reading in the middle class*. Boston, MA: Addison-Wesless.
- Basaraba, D., Yovanoff, P., Alonzo., J., & Tindal, G. (2013). Examining the structure of reading comprehension: *Do literal, inferential, and evaluative comprehension truly exist? Reading and writing*, 26(3).
- Best, J., & Kahan, V. (2003). *Research in education*. New Delhi: Prentice Hall of India.
- Biruk Abera. (2013). *Assessing teacher and student related factors that help improve students' reading comprehension for academic purpose*. (Unpublished M.A Thesis). Haramaya University.
- Brown, H.D. (2004). *Language assessment principles and classroom practices*. New York: Longman.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Carroll, J.B. (1968). The psychology of language testing in A. Davis (ed.): *Language Testing Symposium: A Psycholinguistic Approach*. London: Oxford University Press.
- Calkins, L., Montgomery, K., & Santaman, D. (1998). *A teacher's guide to standardized reading tests: Knowledge is power*. Portsmouth, NH: Heinemann.
- Catala, G. Catala, M. Molina, E. & R. (2001). *Evaluation de la Comprension lectora: Pruebas ACL (Assessment of reading comprehension: ACL test)*. Barcelona: Editorial Grao.

- Council of Europe (2011). Manual for Language Test Development and Examining. *Language Policy Division: Strasbourg. ALTE.*
- Cross, E.A & Carney, E. (1950). *Teaching English in high school*. New York: Macmillan publishers Ltd.
- Dembo, Meyron H. (1994). *Applying Educational Psychology (5th edn.)*. New York: Longman Publishing Group.
- Dessalegn Getachew. (2014). *Assessing the causes of anxiety on students' participation in English classroom*. (Unpublished MA Thesis). Haramaya University.
- Ebel, Robert L. (1979). *Essentials of educational measurement (3rd edn.)*. New Jersey: Prantice. Hall Inc.
- FDRE, MOE. (2002 E.C). English for Ethiopia: Grade 12 Student Textbook. India.
- Gronlund, Norman E. (1981). *Measurement and Evaluation in Teaching (4th edn.)*. New York: Macmillan Publishing Co. Ltd.
- Gordon, C.M. & Hanauer, D. (1995). The Interaction between Task and Meaning Construction in EFL Reading Comprehension Tests. *TESOL Quarterly*.29.
- Fairbairn, S. (2007). Facilitating Greater Test Success for English Language Learners. *Practical Assessment Research and Evaluation*, 12(11). Retrieved from <http://www.pareonline.net/getvn.asp?v=12&n=11> Le. On July 12, 2014.
- Hacuta, K. (2009). *Guidelines for the assessment of English Language Learners*. Educational Testing Service: Stanford University.
- Haladya, T.M. & Downing, S.M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.
- Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. *Reading and writing*, 2(2).

- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge University Press.
- ICDR (2002). *Teacher Education Handbook*. Ministry of Education: Addis Ababa.
- JUCDE. (2004). *Language testing and assessment (TEFL 603)*. Jimma, Ethiopia.
- Jimma University CDE. (2010). *Introduction to Statistics (STAT 202)*. Jimma- Ethiopia.
- Kopriva, R. (2000). Ensuring accuracy in testing for English language learners. Washington, DC: *Council of Chief State School Officers*. Retrieved from <http://www.ccsso.org/content/pdf/EnsuringAccuracy.pdf> on October 15, 2014.
- Kucuk, F. & Walters, J. (2009). How good is your test? *ELT Journal* 63(4).
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33. Retrieved from <http://www.stat.tutorials.com/SPSSDATA> for files mentioned in this @ TexaSoft. 2008.
- Menken, K. (2005). What are the critical issues in wide scale-assessment of English language learners? Issues Brief No.6. Washington, D.C: *National Clearing house for Bilingual Education*. Retrieved on September 3, 2014 from <http://www.ncele/gwu.edu/ncbebriefs//ib6.ltm>
- Mugunda, O., M. & Mugunda, A.G. (2003). *Research Methods: Qualitative and Quantitative Approaches*. Nairobi, Acts Press.
- MOE (2002). *Introduction to applied educational research (Educ.222)*. Educational Media Agency: Addis Ababa, Ethiopia.
- Neuman, W., L. (2003). Social Research Methods. Qualitative and Ethiop. J. Educ. & Sc.Vol. 6 No.1 September 2010108: *Quantitative Approaches (5th ed.)*. Boston: Allyn and Bacon.

- Onwuegbuzie, A.J., & Leech, N.L. (2006). Linking questions to mixed methods. *Data analysis procedures*. The Qualitative Report, 11(3).
- Park, H. M. (2009). Comparing Group Means: *T-test and One -way ANOVA Using STATA,SAS, R, and SPSS, Working Paper*. The University Information Technology Service (UITS) Center for Statistical and Mathematical Computing, Indian University.
- Paynter, D. Bodrova, E., & Doty, J. (2005). For the love of words: Vocabulary instruction that works. San Francisco: Jossey-Bass.
- Popham, W.J. (2005). *Classroom assessment: What teachers need to know* (4thed.). Boston, MA: Pearson Education.
- Putwain, D. W. (2008). Deconstructing test anxiety. *Emotional and Behavioural Difficulties*, 13, 141-155.
- Rivera, Collum, & Sia (2006). An analysis of state assessment policies regarding the accommodation English language learners. In Rivera & Collum (Eds.), *State assessment policy and practice for English language learners: A national perspective*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Selinger, W. & Shohamy, E., (1998). *Second language research methods* (3rdEd). Oxford: Oxford University Press.
- Sukhia, S.P; Mehrotra, P.V. & Mehtrotra, R.N. (1996). *Elements of educational research* (2ndedn.). New Delhi: Allied Publishers Private Ltd.
- Stevens, Butler & Castellon-Wellington., (2000). Academic Language and Content Assessment: Measuring the process of English language learners (ELLs)(CSE Technical Report 552). Los Angeles, CA: *Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing*. Retrieved from <http://www.cse.ucla.edu/Reports/TR552.pdf> on October 6, 2014.

- Tarekegn Wodajo. (2012). A comparative study of reading comprehension skills between governmental and non-governmental preparatory school students of Jimma: Grade 12 in focus (Unpublished MA Thesis). Jimma University.
- Thissen, D., & Wainer, H. (2001). *Test scoring*. Mahwah, NJ: Erlbaum.
- Thorndike, R.L. & Hagen, E.P. (1977). *Measurement and evaluation in Psychology and education (4th edn.)*. New York: John Wiley and Sons.
- Wallace, C. (2003). *Reading*. Oxford: Oxford University Press.
- Wang, P. (2009). The inter-rater reliability in scoring composition. *English Language Teaching*, Britain, 2(3). Retrieved from <http://www.ccsenet.org/journal.html> October 12, 2014.

8. APPENDIXES

Appendix A

Jimma University

College of Social Sciences and Humanities

Department of English Language and Literature

Questionnaire to be filled by Students

Dear students,

This checklist questionnaire is prepared to gather data for MA thesis in TEFL under the title “Assessment of the construct-irrelevant factors that affect students’ performance in reading comprehension test with the particular reference to Grade 11 students.” Therefore, the objective of this checklist questionnaire is to assess the challenges you face while you are doing the reading comprehension test during mid-exam and final exam in 2014/15 *beyond the ability, skill and knowledge* that the test is intended to measure. Your responses will be kept confidential and used only for research purposes. So you are kindly requested to provide genuine responses.

Thank you very much for your cooperation!

Background information

Name of the school -----

Grade -----

Student code -----

Sex -----

Construct-irrelevant factors of language test performance checklist questionnaire

Directions: In the following list you will find the various construct-irrelevant factors that you may encounter while you are doing a reading comprehension test in English classes. Read each of them and put a mark (“x”) in the box under “yes” if you faced a challenge on the test and under “no” if you don’t face a challenge. Then, write your own suggestions for the other challenges you face during reading comprehension test.

No.	Items	Yes	No	Suggestions about the
	<i>The major construct-irrelevant factors you face in doing the reading comprehension test beyond the ability, skill, and knowledge that the test is intended to measure</i>			problems that encounter you when you are doing the test.
1	<p>Are the texts and comprehension questions well-spelt and clear in terms of:</p> <p>a) Spelling</p> <p>b) Punctuation</p> <p>c) Capitalizations</p> <p>If not, what is its adverse effect on your performance? _____</p>			
2	<p>Is the exam paper neat and attractive?</p> <p>If not, what is its consequence? _____</p>			
3	<p>Are the initials of each and every paragraph of the reading passages assigned with numbers?</p> <p>If not assigned with numbers, what could be its influence in doing the test?</p> <p>_____</p>			
4	<p>Is the vocabulary in context asked for its meaning always written in bold letters?</p> <p>What could be the disadvantage of vocabularies not written in bold letters in the given passages? _____</p>			

5	<p>In multiple-choice items of test-question, are the questions and alternative choices sometimes provided on different pages of papers?</p> <p>If so, what is its problem in doing the test? _____</p>			
6	<p>Are the alternatives provided in the form of:</p>			
	<p>1. A._ B._ C._ D._ (Horizontally)</p>			
	<p>In column like: 2. A._ C._ B._ D._</p>			
	<p>3. A. __ B. __ C. __ (vertically) D. __</p>			
	<p>What is the disadvantage of putting the alternatives horizontally? _____</p>			
7	<p>Is correction orientation of test-questions and reading passages given by:</p>			
	<p>a. Invigilator during the test?</p>			
	<p>b. Subject teacher during the test?</p>			
	<p>c. Sometimes by both invigilator and subject-teacher</p>			
	<p>If not, what is its adverse effect? _____</p>			
8	<p>Is disruptive behavior sometimes revealed from the invigilator? If so, what is its adverse effect on test performance? _____</p>			

9	<p>Have you experienced open-ended reading comprehension test in English classes?</p> <p>If not, which type of items you are familiar with in reading comprehension test? _____ Why?</p>			
10	<p>Are the test questions always provided with difficult vocabularies?</p> <p>If not what is its adverse effect on your performance? _____</p>			
	Adverse effects			
11	<p>Do you feel worry for your test scores due to your unfamiliarity with the test format?</p> <p>Why? _____</p>			
12	<p>Do you feel angry so that you hate the subject teacher?</p> <p>Why? _____</p>			
13	<p>Do you feel panic when you see the bad typing format used in texts (reading passages) and questions used in tests?</p>			
14	<p>Do you feel fear when you read difficult vocabularies used in questions during the test?</p> <p>Why? _____</p>			
15	<p>Do you feel test anxious (fear of test) when taking the reading comprehension test?</p> <p>What makes you feel test-anxious?</p>			
	<i>Strategies to overcome the construct irrelevant factors</i>			
16	<p>Usage of careful correction of the passage and test questions by subject teachers</p>			
17	<p>Usage of experienced type writer</p>			
18	<p>Usage of plain language in test-questions by the teacher or test-developer</p>			
19	<p>Familiarizing students with open-ended reading comprehension in classroom test.</p>			

Appendix B

Jimmaa Universiitii

Koolleejjii Saayinsii Hawaasaa fi Himaaniitii

Muummee Afaan Ingilizii fi Hog-barruu

Gaaffilee Cheekliistiin Barattootaaf Qophaa'e

Kaayyoon cheekliistii odeeffannoo kanaa qormaata dubbisa keessaa ba'an deebisuu (Reading comprehension test) irratti danqaalee barattootarratti geessissan jechuunis yeroo qormaataarratti dandeettii, beekumsaa fi ogummaa barataan alatti qabxii barattootaarratti miidhaa waantota fidan adda baasuu dha.

Galatoomaa!

Odeeffannoo Seenaa

Maqaa mana barumsaa: _____

Kutaa: _____

Koodii Barataa: _____

Saala: _____

Qajeelfama: Gaaffilee armaan gadi-jiran irratti wantoota qormaata dubbisa keesaa ba'an irratti danqaalee sitti ta'aniif "eyyee" jalatti mallattoo "√" kaa'i; miti yoo ta'e immoo "lakkii" jalatti mallattoo "√" " kaa'i. Sababa kana jetteefis yaadaakee iddoo kenname irratti gabaabsi barreessi.

Lakk.	Wantootaa	Eyyee	Lakki	Yaada dhimmoota kana ilaalchisee
	<i>Rakkoolee qormaata dubbisa keessaa ba'an (reading comprehension test) irratti dandeettii, ogummaa, fi beekumsa barataan alatti wantoota dhiibbaa geessisan</i>			
1	Dubbisni dhiyaatee fi gaaffileen achi-keessaa ba'an sirnaan kan barra'aniif fi ijattis sirnaan mul'atan:			
	a. Qubeeleen			
	b. Qub-guddeessi			
	c. Sirni-tuqaalee			
	Yoo sirnaan hinbarreffamne ta'e, rakkoo inni qormaaticha hojjechuukeerratti fidu maal faa ta'uu danda'a? Kan sadeen isaanii tokko tokkoon ibsi.			
2	Waraqaan qormaataa dhiyaatu qulqulluu fi hawwataadhaa? Miti yoo ta'e dhiibbaa maalii fidaa? _____			
3	Tokko tokkoon jalqaba keeyyata dubbisichaa lakkoofsaan taa'eeraa? Lakkoofsaan taa'uu dhabuun keeyyataa dubbisa keessatti dhiibbaa maalii fida?			
4	Jechaa fi gaaleen dubbisa keessatti hiikni isaanii gaafataman guddatanii yookaan jala muramanii mul'atuu? Yoo jala hin muranme yookaan guddatanii hin mul'atan ta'e rakkoo maalii fidu jettee yaadda? _____			

5	<p>Gaaffileen qormaataa fi filannoowwan kan deebiin keessaa filataman adda ba’anii fuula adda addaa irratti dhiyaataniiruu?</p> <p>Yoo dhiyaatan rakkoo maalii fiduu?</p>			
6	<p>Filannoowwan gaaffilee deebiin keessaa filatamu akkamiin ta’aniiruu?</p>			
	<p>1. A. __ B. __ C. __ D. __ (dalgeen)</p>			
	<p>Irraan gadee garee-lamaan (column)</p> <p>2. A. __ C. __ B. __ D. __</p>			
	<p>3. A. __ B. __ (irraan-gadee) C. __ D. __</p>			
	<p>Filannon kun dalgeen yoo taa’e miidhaa inni fidu maalidha?</p>			
7	<p>Sirreessuun barreeffama dubbisaa fi gaaffilee dubbisa keessatti doggoggora mul’ataniif sirreeffama kan kennu:</p>			
	<p>a. Qoraa</p>			
	<p>b. Abbaa gosa barnootichaa</p>			
	<p>c. Takka-takka lameen isaaniin</p>			
	<p>Hin kennan yoo ta’e rakkoo maalii fida?</p>			
8	<p>Amalli duddutuu barsiisaa qoru irraa mul’ataa?</p> <p>Yoo mul’ata ta’e qormaata irratti rakkoo maalii fida?</p>			

9	<p>Qormaata dubbisanii deebisuu (reading comprehension test) filannodhaan alatti deebii gabaabaa kennuun qoramtee beektaa?</p> <p>Yoo gaaffileen akkasiin dhiyaatan qoramtee hin beektu ta'e, guyyaa gaafa qormaata akkasii fudhatu itti wallaaluun dhiibbaa akkamii sirratti uuma? Akkamiin?</p>			
10	<p>Gaaffileen dubbisichaaf dhiyaatan jechoota yookaan gaalee hiikni isaanii ulfaatuun dhiyaataniiruu?</p> <p>Dhiyaataniiru yoo ta'e rakkoon inni qormaata hojjechuukee irratti fidu maali dha?</p>			
	Dhiibbaalee mul'atan			
11	<p>Qabxiin najalaa xiqqaata yookaan nankufa jettee ni yaaddoftaa?</p> <p>Maaliif?</p>			
12	<p>Aaruurraan kan ka'e gosa barnootichaa fi barsiisaa gosa barnootichaa barsiisurratti jibbinsa horattaa?</p> <p>Maaliif?</p>			
13	<p>Jalqabuma akkaataa barreffamaa ilaaluun qormaata irratti nijeeqamtaa?</p> <p>Maaliif?</p>			
14	<p>Gaalee fi jechoota baay'ee ulfaataa ta'an ilaaluun qormaata irratti ni jeeqamtaa?</p> <p>Maaltu ta'a jettee jeeqamta?</p>			

15	Yeroo qormaatni dhiyaatu sodaan qormaataa sitti dhaga'amaa? Kun immoo dhiibbaa maalii fida jettee yaadda?			
	Tooftaalee dhiibbaa kun itti hiir'ifamu			
16	Nama sirnaan ogummaa komputaraa qabuun fayyadamuun barreessuu			
17	Barsiisaan gosa barnootichaa barsiisu otuu qormaatni barattootaaf hin dhiyaatiin dura sirnaan sakatta'uun sirreessuu			
18	Barsiisaan qormaata baasu gaaffilee keessatti jechoota sasalphaa gargaaramuun eggannoodhaan qormaata baasuu			
19	Deebii gabaabaa kennuu daree keessatti qoramu shaakaluu.			

Appendix C

Jimma University

College of Social Sciences and Humanities

Department of English Language and Literature

Reading Comprehension Test for Students

Dear Students,

This reading comprehension test is prepared to gather data for MA thesis in TEFL under the title “Assessment of the Construct-irrelevant Factors Affecting Students Reading Comprehension test with particular reference to Grade 11.” The objective of this test is to gather the language related factors that affect student’s performance in reading comprehension test with the four modes of writing (descriptive, expository, argumentative, and narrative) beyond the ability, skill, and knowledge that the test is intended to measure. Thus, your genuine response to the questions taken from the paragraphs will have significant contribution for the success of this study.

Thank you very much for your cooperation!

Session one

Write short answers for the questions taken from paragraphs on the space provided.

Text One

My dorm room has a very good location in the women’s dormitory building. It is located on the second floor, which means that I do not have to climb too many stairs. I can go down to the study and the canteen, then up to my room again easily. When I forget something or leave my books in the room it is no problem to back to the room. I also feel safe on the second floor. It is not on the ground floor, and I do not have to worry about the thieves climbing into my room through the window, and free of noise disturbances. In short, my room has an excellent location and I think I am lucky to have a room like that.

1. What is the crux or essence of this paragraph?

2. What are the downsides of living on the ground floor according to the above paragraph?

a. _____

b. _____

Text Two

Air pollution is one of the causes of climatic changes in different parts of the world. If the pollution affects the level of carbon dioxide in the atmosphere, the results are likely to be serious. Carbon dioxide constitutes only small part of the atmosphere. But it has an important function in maintaining the balance between radiation from the sun entering the atmosphere and radiation leaving the Earth. Some of the radiation is absorbed by the Earth and some is radiated back into the atmosphere. The carbon dioxide in the atmosphere prevents some of the radiation from leaving the atmosphere. Thus, the air remains in the atmosphere and carbon dioxide helps to prevent the atmosphere of the Earth from falling.

3. What are the pros and cons of carbon dioxide according to the given paragraph?

a. _____

b. _____

Text Three

Some people argue that foreign tourists should be encouraged to use local transport to visit sites. They argued that local transport is cheap and that is the best way to experience local culture. However, my own view is that this is inadvisable for several reasons. Firstly, since tourists are unlikely to speak the local language, they will not be able to communicate with other passengers. Another problem is that they may be target for thieves. In addition, local buses are often crowded and uncomfortable. I would argue that tourists should use a four wheel drive vehicle hired from a tour company in order to visit tourist sites. This would allow them to travel in reliable transport, accompanied by a driver, who would also act as a guide and interpreter. Thus, they would be able to visit any tourist site safely, comfortably and without problem of communication.

4. What are the justifications that the tourists shouldn't be inspired to use local transport?

a. _____

b. _____

c. _____

5. What about the justification that tourists should be prompted to use local transport?

a. _____

Text Four

There was a new girl, Ginny, in Josie's class who wore really thick glasses, and without them, was nearly blind. She, to my relief, had temporarily become the butt of jokes and pranks. The latest chant that the kids had come up with was, "Ginny, Ginny, short and fat, squinty-eyed and blind as a bat!" In all fairness, Ginny wasn't fat at all, but the kids chanted that because it rhymed with bat.

6. Provide a cogent reason why the kids chanted "Ginny, Ginny, short and fat, squinty-eyed and blind as a bat?"

a. _____

Appendix D

Jimma University

College of Social Sciences and Humanities

Department of English Language and Literature

Reading Comprehension Test for Students

Dear Students,

This reading comprehension test is prepared to gather data for MA thesis in TEFL under the title “Assessment of the Construct-irrelevant Factors Affecting Students Reading Comprehension Test with Particular Reference to Grade 11.” The objective of this test is to gather the language related factors that affect students performance in reading comprehension test using the four modes of writing (descriptive, expository, argumentative, and narrative). Therefore, your genuine response to the questions taken from each paragraphs will have a significant contribution for the success of this study.

Thank you very much for your cooperation!

Second Session

Text One

Write short answers for the questions taken from the paragraphs on the space provided.

My dorm room has a very good location in the women’s dormitory building. It is located on the second floor, which means that I do not have to climb too many stairs. I can go down to the study and the canteen, then up to my room again easily. When I forget something or leave my books in the room it is no problem to back to the room. I also feel safe on the second floor. It is not on the ground floor, and I do not have to worry about the thieves climbing into my room through the window, and free of noise disturbances. In short, my room has an excellent location and I think I am lucky to have a room like that.

1. What is the main idea of this paragraph?

2. What are the disadvantages of living on the ground floor according to the above paragraph?

a. _____

b. _____

Text Two

Air pollution is one of the causes of climatic changes in different parts of the world. If the pollution affects the level of carbon dioxide in the atmosphere, the results are likely to be serious. Carbon dioxide constitutes only small part of the atmosphere. But it has an important function in maintaining the balance between radiation from the sun entering the atmosphere and radiation leaving the Earth. Some of the radiation is absorbed by the Earth and some is radiated back into the atmosphere. The carbon dioxide in the atmosphere prevents some of the radiation from leaving the atmosphere. Thus, the air remains in the atmosphere and carbon dioxide helps to prevent the atmosphere of the Earth from falling.

3. What are the advantage and disadvantage of carbon dioxide according to the given passage?

a. _____

b. _____

Text Three

Some people argue that foreign tourists should be encouraged to use local transport to visit sites. They argued that local transport is cheap and that is the best way to experience local culture. However, my own view is that this is inadvisable for several reasons. Firstly, since tourists are unlikely to speak the local language, they will not be able to communicate with other passengers. Another problem is that they may be target for thieves. In addition, local buses are often crowded and uncomfortable. I would argue that tourists should use a four wheel drive vehicle hired from a tour company in order to visit tourist sites. This would allow them to travel in reliable transport, accompanied by a driver, who would also act as a guide and interpreter. Thus, they would be able to visit any tourist site safely, comfortably and without problem of communication.

4. What are the reasons that the tourists should not be encouraged to use local transport?

a. _____

b. _____

c. _____

5. What about the reason that tourists should be certain to use local transport?

a. _____

Text Four

There was a new girl, Ginny, in Josie's class who wore really thick glasses, and without them, was nearly blind. She, to my relief, had temporarily become the butt of jokes and pranks. The latest chant that the kids had come up with was, "Ginny, Ginny, short and fat, squinty-eyed and blind as a bat!" In all fairness, Ginny wasn't fat at all, but the kids chanted that because it rhymed with bat.

6. Provide convincing reason why the kids chanted "Ginny, Ginny, short and fat, squinty-eyed and blind as bat?"

a. _____

Appendix E:

Jimma University

College of Social Sciences and Humanities

Department of English Language and Literature

Semi-structured Interview for Teachers

Dear teacher,

This interview question is prepared to gather data for MA thesis in TEFL under the title: “Assessment of the construct-irrelevant factors that influence students’ performance in reading English comprehension test.” The objective of this interview is to gather information and to examine the factors that influence students’ performance in reading comprehension test you might have noticed beyond the ability that the test is intended to measure. Thus, your genuine response to the questions will have significant contribution for the success of this study.

Thank you very much for your cooperation!

Background information

1. Name of school _____
2. Teacher code _____
3. Sex _____
4. Teaching experience _____

Construct-irrelevant factors of language test performance based interview

The questions are answered based on the construct-irrelevant factors that affect students’ performance while they encountered on reading comprehension test beyond the ability, skill, and knowledge that the test is intended to measure.

- Q1. What do you think are the major factors influencing students’ performance in reading comprehension test beyond the ability, skill, and knowledge that the test is intended to measure in English classes?
- Q2. What kinds of situations and test-taking activities have you found to be lowering students’ performance while they are doing the test?

- Q3. What are the adverse effect of the construct-irrelevant factors that influence students' performance in reading comprehension test?
- Q4. How do you think the construct-irrelevant factors influencing students' performance can be successfully alleviated in reading comprehension test?
- Q5. Which questions/test items are students more familiar with in English language test?
A multiple-choice item or short-answer item? Why?
- Q6. What are the most commonly used words or phrases or sentences that are used to ask questions from the reading passage? Is there a difference among test language, language of instruction, and everyday language in terms of constructing reading comprehension test?

Thank You!

Appendix F

Jimma University

College of Social Sciences and Humanities

Department of English Language and Literature

Document Analysis (From Exam Bank)

English reading comprehension test document analysis from the exam-bank prepared by teachers for mid-exam and final exam in 2014/15 at Sire and Ano preparatory schools of Grade 11 by the researcher.

Background Information

School _____	Semester _____
Grade _____	Year _____
Topic of the text _____	Mid-Exam _____
Item format _____	Final-Exam _____
Number of questions _____	weighting _____

1. How are the layout and legibility of the text and test-questions of the given reading comprehension test?
2. Is each paragraph of a given text assigned with numbers?
3. Is a paragraph from a paragraph well-spaced from one another?
4. What types of languages (test languages, instructional languages, and everyday languages) are used to ask questions from the given reading passage?
5. Are a text and a comprehension questions presented on different pages of papers?

APPENDIX G

Percentages of Student Responses with Different Construct-irrelevant Variables

Table 1: Matters of Language

No.	Questions	Yes	%	No	%
	Language Related Factors				
10	Are the test questions sometimes provided with difficult vocabularies?	37	84	7	16
	Adverse effects of language related factors				
14	Do you feel panic when you read the difficult vocabularies used in questions during reading comprehension test?	38	84.4	6	13.6
	Strategies to overcome language related factors				
18	Usage of plain language in test questions by the test developers/teachers.	41	93.2	3	6.8

Table 2: Ineffective use of test accommodations

No.	Questions	Yes	%	No	%
	Test accommodation related factors				
7	Is the correction orientation of test questions and texts (reading passages) given by:				
	A: invigilator during the test?	16	36.4	10	22.7
	B: subject-teacher during the test?	3	6.8	7	15.9
	C: sometimes by both the invigilator and subject teacher?	3	6.8	5	11.4
8	Is disruptive behavior sometimes revealed from invigilator?	37	84	7	16
	Adverse effects of test accommodation factors				
12	Do you feel angry so that you hate the subject teacher?	35	79.5	9	20.5
	Strategies to overcome test accommodation factors				
16	Careful correction of texts and test questions by the subject teacher prior test administration?	42	95.5	2	4.5

Table 3: Unfamiliarity of students with item/test format

No.	Questions	Yes	%	No	%
	Item/test format factors				
9	Have you experienced the open-ended reading comprehension test?	4	9	40	91
	Adverse effects of item/test format factors				
11	Do you feel worry for your test score due to your unfamiliarity with the test format?	35	80	9	20
	Strategies to overcome item/test format factors				
19	Familiarize students with the short-answer items through practicing reading comprehension test?	41	93.1	3	6.9

Table 4: The badly typed texts and questions

No.	Questions	Yes	%	No	%
	The badly typed texts and question factors				
1	Are the texts (reading passages) and comprehension questions well-spelt and clear in terms:				
	a. Spelling	19	43.2	5	15.4
	b. Capitalization	8	18.2	4	9.1
	c. Punctuation	6	13.6	2	4.5
2	Is the exam paper neat and attractive?	6	14	38	86
3	Are the initials of each and every paragraphs of the passage assigned with numbers?	3	7	41	93
4	Is the vocabulary in the text (passage) asked for its meaning always written and provided in bold letters?	13	30	31	70
5	In multiple-choice items of test-questions, are the questions and alternative choices always provided on the same pages of papers?	12	27	32	73
6	Are the alternative-choices provided in the form of:				
	1.A._ B._ C._ D._ (or Horizontally)	20	45.5	7	15.9
	2.A._ C._ (or in Column) B._ D._	4	9.1	6	13.6
	3.A._ B._ C._ (or Vertically) D._	4	9.1	3	6.8
	Adverse effects of the badly typed texts and test-questions				
13	Do you feel tension when you review the bad typing format of the test?	39	88.6	5	11.4

15	Do you feel test-anxious (fear of test) during the reading comprehension test?	41	93.2	3	6.8
	Strategies to overcome the badly typed texts and test questions				
17	Usage of well-experienced and computer skilled persons in writing the reading comprehension test.	37	84	7	16

APPENDIX: H

Students' Test Scores in Reading Comprehension Test into Two Sessions

St. code	T1	T2	T3	T4
1	63	85	60	85
2	40	85	45	85
3	50	75	45	70
4	60	85	60	80
5	85	95	87	98
6	55	85	50	85
7	70	95	65	94
8	15	45	10	50
9	40	70	35	65
10	60	60	55	60
11	50	95	55	98
12	35	65	30	63
13	30	45	30	45
14	23	50	20	55
15	50	70	45	65
16	35	70	33	55
17	40	60	38	55
18	25	30	25	35
19	30	40	30	40
20	35	50	34	45
21	20	35	15	40
22	40	65	35	60
23	50	60	55	60
24	50	55	50	55
25	20	50	20	50
26	60	65	60	65
27	35	40	33	40
28	55	55	52	55
29	20	50	23	55
30	25	35	25	30
31	15	10	13	15
32	40	50	42	55

33	10	20	10	20
34	55	65	55	70
35	50	50	50	45
36	40	40	45	40
37	40	40	35	40
38	12	25	10	25
39	55	60	50	65
40	40	40	40	40
41	37	50	35	45
42	22	45	20	45
43	35	45	35	50
44	55	55	55	55

KEY: St .code=students' code, T1=test with unfamiliar language, T2=test with familiar language, T3=test with unfamiliar language, T4=test with familiar language.

APPENDIX I

The Interpretation of Proportions with Equivalent Percentages

For the interpretation and discussion of students' suggestion from the checklist questionnaire, the equivalent percentages for the given proportions were taken from the following lists.

Less than half = 45%

More than a third = 40%

The majority = 75%

A small minority = 10%

The small majority = 60%

The vast majority = 90%

Just under four fifths = 77%

About a half = 52%

All most all/everyone = 95%

A very small number = 5%

Less than one person per hundred = 0.5%

(Source: Grade 12 English Textbook, p.125)

APPENDIX J

The Sample Transcription of Student Suggestion from Checklist Questionnaire

Code 01 student

1. Are the text (passages) and comprehension questions well-spelt and clear in terms of?

a. Spelling

There are words which are spelt wrongly and this exposes learners to give wrong answers due to misinterpretation.

b. Capitalization

Sometimes the proper nouns, and mountains and days of the week are not written in capital letters in the reading passages presented for test-takers.

c. Punctuation

Most sentences are not provided with correct punctuation. For example, the absence of comma between sentences. Therefore, the absence of comma results in difficulties to identify a sentence from a sentence.

2. Is the exam-paper neat and attractive?

The paper is full of texts written in a concentrated way with an inappropriate marginalization from the left and right side of the paper. Thus, it results in invisibility which students cannot understand it. Therefore, this incident affects students test score.

3. Is each paragraphs of the text (reading passage) assigned with numbers?

Totally each paragraphs of the text is not assigned with numbers. Therefore, it kills test-takers time in finding the information from the paragraph because students encounter another layer of difficulty. I.e. counting the paragraphs.

4. Is vocabulary in the text (reading passages) always written and provided in bold letters?

Sometimes vocabularies in context from the passage are written in italics and sometimes not written in bold letters. Therefore, it cannot be easily accessible and visible. This in turn affects student's time in finding it from the text.

5. In multiple-choice items of test questions, are the questions and alternative choices always provided on the same pages of papers?

Sometimes the questions and alternative choices are presented on different pages of papers. This confuses the students in that which questions stands for which alternative choice and vice-versa. It consumes student's time in searching questions and answers from different pages.

6. Are the alternative choices provided in the form of?

1. A_ B_ C_ D_ or (in Horizontal way)

2. A_ C_ or (in Column)

B_ D_

3. A_

B_ or (Vertically)

C_

D_

The alternative choices are presented mostly in a horizontal way with the same line of the test questions which is relatively difficult to identify easily. Therefore, it is advisable if the alternative choices are written vertically in the form of choice 3 in a reading comprehension test.

7. Are correction orientation of test questions and texts or reading passages given by:

a. invigilator during the test?

b. subject-teacher during the test?

c. sometimes by both invigilator and subject-teacher?

Mostly correction-orientation of test-questions and texts are given by the invigilator guided by the subject-teacher. Thus, he/she is not efficient like that of subject teacher. Therefore, the students raise questions again and again and this results in conflict with invigilator. It is therefore, this situation affects student's time and test-scores.

8. Is disruptive behavior sometimes revealed from invigilator?

The disruptive behavior revealed from the invigilator during the test caused conflict between students and invigilator so that it results in tension from the side of students.

9. Have you experienced open-ended reading comprehension test?

I experienced multiple-choice items on reading comprehension test. When I encountered open-ended reading comprehension test, may be, I find it difficult to reach at the answers quickly.

10. Are the questions sometimes provided with difficult vocabularies?

Sometimes I encountered new vocabularies and expressions which are difficult to answer it. As a result, I feel nervous and think about my score.

11. Do you feel worry for your test score due to unfamiliarity with the test format?

I never feel worry for my test score due to unfamiliarity with the test because even if I am unfamiliar with the test format, it will help me to experience such items.

12. Do you feel angry so that you hate the subject teacher due to time constraint you faced to do the test?

The ineffective time adjustments allotted for the test by the subject teacher creates feeling under pressure about the time and the ineffective guidance from the subject teacher during the test creates emotional conflicts and tension that leads to the hatred of subject teachers.

13. Do you feel tension when you review the bad typing format of the test?

When I saw the bad typing format of the text and test-questions during the test, I feel a high level of anxiety that leads to fear and loss of ability to function with an appropriate manner.

14. Do you feel panic when you review the difficult vocabularies used in questions during the reading comprehension test?

I feel panic because the questions are unclear and not timely to answer due to difficult vocabularies.

APPENDIX K

Sample Transcription of Teachers' Interview

Teacher A

1. What do you think are the major construct-irrelevant factors affecting students' performance in reading comprehension test? i.e. beyond the ability, skill and knowledge that the test is intended to measure?

I think there are many factors that affect students' language test performance in reading comprehension test beyond the ability, skill and knowledge that the test is intended to measure. To name some of them, complex sentences, ambiguous terms or tricky languages, inappropriate time allocation, vague instructions which is not highly structured and stated, the way the exam is written, seating arrangement, the careless test construction of teachers, some learners and teachers disruptive behaviors are some of the major ones that affect students test performance.

2. What kinds of situations and test-taking activities have you found to be lowering students performance while they are doing the test?

Here, the inappropriate seating arrangement paves the way for cheating and this cheating result in an intimidation from the invigilator which in turn creates conflict between students and invigilator is one of the situation and test-taking activities that affects students' test performance.

3. What are the adverse effects of the construct-irrelevant factors on student's performance in reading comprehension test?

I think the construct-irrelevant factors have their own adverse effects on students' test performance particularly in reading comprehension test. For example, loss of interest on the side of students that leads them to guessing the answers for the questions taken from the passage without reading it and jump over it to do the items such as multiple-choice items.

4. How do you think that the construct-irrelevant factors influencing students' performance can be successfully alleviated in reading English comprehension test?

I think the subject teacher can carefully prepare the test by giving attention to the tests clear instructions, balanced time allotment, mechanics, usage of easily accessible language to the students in test questions and finally the provision and allocation of abundant test materials by school administrative bodies and before the beginning of the exam, clear instructions should be given to test-takers about how to behave during the examination, invigilators should know what to do if regulations are broken, or there are other events, foreseen or unseen are some of the strategies used to alleviate the mentioned problems. And also the language department should create awareness on how to construct test free of construct-irrelevant factors with the support of nearby universities.

5. Which test items/questions students are more familiar with in English reading comprehension test?

In our school, students are more familiar with multiple-choice items in classroom tests. The reason is that it is easily to score within a short of time objectively by the teacher.

6. What are the most commonly used words/phrases/sentences that are used to ask questions from the reading passage? Is there a difference among language test, language of instruction and everyday language when constructing English classroom tests?

I think the most commonly used words include topic, infer, refer/s, conclude etc and the most commonly used phrases to construct tests are according to, the expression/s, main idea, the word are used to ask questions from the reading passage. On the sentence level, for example,

What do you think the main idea of this passage?

What is the topic of this passage?

What would you infer from the given paragraph?

Which one of the following is not mentioned in paragraph one?

The expression “_____” in line _____ refers to _____.

On the other hand, words like “discuss” and “explain” can be ambiguous and are not the test languages when they are not given with specific instructions as to what points should be discussed to ask questions from the reading comprehension test.

APPENDIX L

Sample Transcription of Exam Paper Analysis

School: Ano Preparatory School

Grade: 11

Topic of the test: The fortified historic town of Harare

Item format: Multiple- choice

Number of questions: 8

Semester: I

Year: 2014/15

Test prepared for: final-exam

Weighting: 2 points each

1. Does the text provided and test-questions well laid and perfectly legible?

The text entitled “The fortified historic town of Harare” is a descriptive essay presented for test-takers in Ano Preparatory School of Grade 11. The questions are presented in the form of multiple-choice items weighting two points each. It is a final exam in the first semester of 2014/2015. Therefore, the given text is not well laid out and perfectly legible. There is incorrect spelling, wrong punctuations and the usage of too much text in too a small a space due to the minimized size of letters. Also the questions and alternative choices are written continuously and horizontally in the same line.

2. Is each paragraph of a given text assigned with numbers?

The text entitled “The fortified historic town of Harare” has eight paragraphs and all of the paragraphs are not assigned with numbers.

3. Is a paragraph from a paragraph well spaced?

The text entitled “The fortified historic town of Harare” has eight paragraphs which are written with full passages that do not attract the test-takers.

4. Is the test languages used in test questions?

The text entitled “The fortified historic town of Harare” is followed by 8 multiple choice items. Out of these questions, one question number 3 is presented with difficult languages (vocabularies). The question is presented in the following way: if there are magnificent mosques and also Christian churches, we infer that there are both the laity of: in such question, **laity** is a difficult word for the test taker to answer the question since the word “laity” is not in context.

5. Is a comprehension questions and alternative choices presented on different pages of papers?

The question taken from the passage, for example, in question number 6, the question and the alternative choices (A-E) are presented on separate pages of papers. In relation to this, the alternative choices presented in the form of “all of the above”, “none of the above”, A and B are against the rule of Brown’s law that explains “If one distracter is ridiculous, that distracter is not helping to test students. Instead, those students who are guessing will be able to dismiss that detractor and improve their chances of answering the item correctly without really knowing the correct answer (JUCDE, 2014). Most teachers want their tests to be relatively efficient by adding extra-information that is not related to the construct being tested. This will just take extra-time for students to read and will add nothing to the test. Such extra-information may also advertently provide students with clues that they can use in answering other items.