

The Practice of English Language Test Items Construction and Administration in Second Cycle Primary Schools

By
Esayas Bekele Mekete

A Thesis Submitted to Adama Science and Technology University
School of Humanities and Law
Department of English

September, 2017
Adama, Ethiopia



ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY
SCHOOL OF HUMANITIES AND LAW
DEPARTMENT OF ENGLISH

**The Practice of English Language Test Items Construction
and Administration in Second Cycle Primary Schools**

By
Esayas Bekele Mekete

Advisor
Teshome Tola (PhD)

**A Thesis Submitted to Adama Science and Technology University
in Partial Fulfillment of the Requirement for the Degree of Masters in
Teaching English as a Foreign Language (TEFL)**

September, 2017
Adama, Ethiopia



DECLARATION

I hereby declare this thesis on the title “**The Practice of English Language Test Items Construction and Administration in Second Cycle Primary Schools**” is my original work and that all sources used for the thesis have been indicated and acknowledged with complete references. It has not also been submitted for any assessment at Adama Science and Technology or any other university or institution.

Name: **Esayas Bekele Mekete**

Signature: _____

Date: _____

Place: Adama Science and Technology University School of Humanities Department of English

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF HUMANITIES AND LAW

DEPARTMENT OF ENGLISH

THE PRACTICE OF ENGLISH LANGUAGE TEST ITEMS CONSTRUCTION AND
ADMINISTRATION IN SECOND CYCLE PRIMARY SCHOOLS

BY

ESAYAS BEKELE MEKETE

APPROVAL OF THE EXAMINATION BOARD

NAME

SIGNATURE

DATE

Dr. Teshome Tola



01/10/2017

ADVISOR

Tamene Kiths

Tamene 

01/10/2017

EXTERNAL EXAMINER

Alemu Disassa (PhD)



01/10/2017

INTERNAL EXAMINER

Acknowledgements

Firstly, my heartfelt gratitude goes to my thesis advisor Dr. Teshome Tola for his prudent and constructive comments on my research paper from the beginning to the end. Then, my special thank is credited to my colleagues who helped me vet the data gathering tools, participated in the English language classroom observations and evaluation of English language test papers. I am also thankful to English language teachers and students of Abomsa number 1 and 2 schools who took part in the pilot study.

Moreover, I would like to thank English language teachers, language department heads and students in Jeju Woreda who participated in the study by giving information for the success of the study.

In addition, my deepest gratitude goes to my friend Samuel Abera who helped me proofread the thesis. I am also grateful to the school directors and clustered resource center (Henceforth, CRC) supervisors of second cycle primary schools (Henceforth, SCPS) and Ato Merete Seifu, who is the officer of curriculum and examinations in Jeju Woreda education office, for their cooperation during the distribution and collection of the questionnaire.

Furthermore, I am indebted to Ato Dereje Lema who translated the students' questionnaire into Afan Oromo.

Last but not least, I extend my thanks to my family, my school director Ato Shimellis Dereje and workers of World Vision Ethiopia Jeju area development program for their moral and material support.

Table of Contents

Contents	Pages
Acknowledgements.....	i
Table of Contents.....	ii
List of Tables	vi
List of Figures.....	vii
List of Appendices	viii
Acronyms.....	ix
Abstract.....	x
Chapter One: Introduction	1
1.1 Background of the Study	1
1.2 Statement of the Problem	2
1.3 Objectives of the Study	4
1.3.1 General Objective of the Study	4
1.3.2 Specific Objectives of the Study	4
1.4 Research Questions	5
1.5 Significance of the Study.....	5
1.6 Scope of the Study.....	5
1.7 Limitations of the Study	5
1.8 Operational Definitions of Terms.....	6
1.9 Organization of the Study.....	6
Chapter Two: Review of Related Literature	7
2.1 What is a Test?	7
2.2 Perspectives of Testing.....	7
2.3 Why a Test?.....	8

2.4	Types of Tests.....	8
2.5	Teaching and Testing	9
2.6	The Focus of Language Tests.....	9
2.7	Qualities of a Good Test.....	9
2.7.1	Validity.....	10
2.7.2	Reliability	11
2.7.3	Authenticity	12
2.7.4	Practicality.....	12
2.7.5	Washback	12
2.7.6	Fairness	13
2.8	Essentials in Test Item Construction	13
2.8.1	The Role of Learning Objectives in Test Item Construction	13
2.8.2	The Role of Table of Specification	14
2.8.3	Test Item Formats.....	15
2.8.3.1	Multiple Choice Items.....	16
2.8.3.2	True-False Items	18
2.8.3.3	Matching Items	19
2.8.3.4	Short Answer Items.....	20
2.8.3.5	Essay Test Items	21
2.9	General Phases of Test Items Construction.....	22
2.10	Drawbacks in Teacher-made Tests.....	25
2.11	Test Administration	28
2.11.1	Test Preparation Considerations.....	29
2.11.2	Psychological Aspects.....	30
2.11.3	Physical and Environmental Conditions	31
2.11.4	Time and Place Considerations	31

2.11.5 Considerations of Extraneous Variables	31
Chapter Three: Research Design and Methodology	33
3.1 The Research Design	33
3.2 Research Setting	33
3.3 Population , Samples and Sampling Technique	34
3.4 Data Collecting Instruments	35
3.4.1 Questionnaires	35
3.4.2 Document Analysis	36
3.4.3 Semi-structured Interview	36
3.4.4 Class Observation.....	36
3.5 The Pilot Study	37
3.5.1 Reliability of the Questionnaires	37
3.5.2 Validity of Data Collecting Instruments	37
3.6 Data Collection Procedures	38
3.7 Data Analysis and Interpretation Procedures	39
3.8 Ethical Considerations.....	40
Chapter Four: Presentation, Analysis and Interpretation of the Data	42
4.1 Demographic Characteristics of the Respondents	42
4.1.1 Sex.....	43
4.1.2 Age	44
4.1.3 Academic Status of Teachers	44
4.1.4 Academic Field of Teachers.....	44
4.1.5 Teaching Experience of Teachers	44
4.2 Teachers' Response to the Application of Principles of Planning Classroom Tests	45
4.3 Teachers' Response to the Application of Principles of Classroom Test Items Construction	48
4.4 Teachers' Response to the Application of the principles of Revising Classroom Tests	52

4.5	Triangulation of Data in Planning, Constructing and Revising Classroom Tests	54
4.6	Teachers' and Students' Response to the Application of Pre-test Administration Principles.....	55
4.7	Teachers' and Students' Response to the Application of While-test Administration Principles	60
4.8	Teachers' and Students' Response to the Application of Post-test Administration Principles	64
4.9	Triangulation of Data in Pre-test, While-test and Post-test Administration	66
Chapter Five: Summary of the Findings, Conclusions and Recommendations.....		67
5.1	Summary.....	67
5.2	Conclusions	70
5.3	Recommendations	71
References.....		73

List of Tables

Table 1: Demographic characteristics of the respondents	43
Table 2: Teachers' response to the extent to which they applied principles of planning classroom tests.....	45
Table 3: Teachers' response to the extent to which they applied the principles of classroom test items construction.....	48
Table 4: Teachers' response to the extent to which they revised classroom tests	52
Table 5: Teachers' response to the extent to which they applied principles of pre-test administration	55
Table 6: Students' response to the extent to which their English teachers applied principles of pre-test administration	56
Table 7: Teachers' response to the extent to which they applied principles of while test administration	60
Table 8: Students' response to the extent to which their English teachers applied principles of while test administration.....	61
Table 9: Teachers' response to the extent to which they applied principles of post-test administration	64
Table 10: Students' response to the extent to which their English teachers applied principles of post-test administration.....	64

List of Figures

Figure 1: Overall mean scores in English language test items construction.....	54
Figure 2: Overall mean scores in English language test items Administration	66

List of Appendices

Appendix A: Questionnaire to Be Filled by English Language Teachers	78
Appendix B: Questionnaire to Be Filled by Grade 7 Students	84
Appendix C: Questionnaire to Be Filled by Grade 7 Students Amharic Version	87
Appendix D: Questionnaire to Be Filled by Grade 7 Students Afan Oromo Version.....	90
Appendix E: Classroom Observation Checklist on Test Administration	93
Appendix F: English language Test items construction Evaluation Checklist	95
Appendix G: Interview Questions for Language Department Heads	98
Appendix H: English Language Test Paper in the First Semester of 2016/17	99
Appendix I: English Language Test Paper in the First Semester of 2016/17	100
Appendix J: English Language Test Paper in the First Semester of 2016/17.....	101
Appendix K: English Language Test Paper in the First Semester of 2016/17	102
Appendix L: English Language Test Paper in the First Semester of 2016/17	104
Appendix M: English Language Test Papers in the Second Semester of 2015/16.....	105
Appendix N: English Language Test Papers in the Second Semester of 2015/16	107
Appendix O: English Language Test Papers in the Second Semester of 2015/16	108
Appendix P: English Language Test Papers in the Second Semester of 2015/16.....	109
Appendix Q: English Language Test Papers in the First Semester of 2015/16.....	110
Appendix R: English Language Test Papers in the First Semester of 2015/16.....	111
Appendix S: English Language Test Papers in the First Semester of 2015/16	112
Appendix T: English Language Test Papers in the First Semester of 2015/16	113

Acronyms

A.M- Anti-Meridian

CELA- Cambridge English Language Assessment

CRC- Cluster Resource Center

ICDR- Institute of Curriculum Development and Research

MoE- Ministry of Education

PM- Prime Meridian

SCPS- Second Cycle Primary School

SD- Standard Deviation

SPSS- Statistical Package for Social Science

TEFL-Teaching English as a Foreign Language

ToS- Table of Specification

Abstract

The main objective of this study was to analyze the practice of English language test items construction and administration in the second cycle primary schools (Henceforth, SCPS) of Jeju Woreda. To achieve this, the study employed a descriptive survey research design. The samples of the study were selected through the combination of census and simple random sampling techniques. Hence, all English language teachers who were teaching in the SCPSs of the Woreda were selected using census sampling, whereas, eight department heads, eighty-eight grade seven students, thirteen teachers-made English language test papers and seven grade seven English language classes were selected using simple random sampling techniques. The instruments used to collect data were the questionnaire, semi-structured interview, document analysis and observation. The data gathered through closed-ended items were coded, tabulated and analyzed using descriptive statistics such as frequency, mean and standard deviation, whereas, the additional data gained from the open-ended items, semi-structured interview, documents (English language test papers) and English class observations were also coded, categorized and identified thematically in order to approve or disapprove the quantitative data. The result of the study showed that English language teachers were aware of the principles of test items construction and administration with mean 2.84 and 3.08 respectively. Contrary to this, the thematic analysis of the qualitative data showed they developed test items without adequate planning in taking no notice of the principles of test items construction. Hence, the result of document analysis depicted that most English language test papers consisted of ambiguous, incomplete, wrong and biased items, and grammatical, spelling and typing errors. English language test items also focused on structure and vocabulary that ignored other language skills. Moreover, the constructed tests were not administered based on the principles as the data from the class observations showed. Generally, English language teachers didn't usually apply the principles of English language test items construction and administration although they seemed to know the principles. Based on the findings, it is recommended that English language teachers should take time to prepare test items considering the principles of test items construction. Contextualized training should also be given to English language teachers on how to construct and administer qualified English language tests on the basis of defects they frequently make. Moreover, colleges of teacher training should train teachers on how to construct and administer tests.

Key terms: construction, administration

Chapter One: Introduction

This section incorporates background of the study, statement of the problem, objectives, research questions, significance, scope, limitation of the study, operational definitions of key terms and organization of the study. These are discussed one by one.

1.1 Background of the Study

No language can be compared with English in respect to its global status. This reveals that it is used to communicate social, economic, educational and political issues. It is also a language of commerce, science and technology. Besides, it's the language of air and maritime navigation in the worldwide network (Bachore, 2015).

In Ethiopia, the English language plays a pivotal role in education. It is a language of trade, business, advertisement, entertainment and administration (ibid). According to the Ethiopian education policy, English is a foreign language learned for the purpose of "... cultural and international relations" with the outside world (MoE, 1994:24). Therefore, it is taught as a subject starting from grade one and serves as a medium of instruction for secondary and higher education. Hence, it should be taught and tested appropriately to achieve its ultimate goal. Pertaining to this, Siddiek (2010) proposes the following:

It is not difficult to teach a unit or course to [---] students, but it is important to know how successful that teaching is. The success of that course or unit can be seen through the right assessment of the efforts exerted to teach the course. Good teaching consequently yields good testing. The good testing is supposed to yield a positive backwash on both the learner and the teacher simultaneously (p.142).

This means that testing is an important part of the teaching and learning process. It is helpful to examine the progress of students in a given educational system. It also helps teachers to address the desired goal of education. In addition, it gives feedback to the teachers whether the teaching methodology they used is effective or not. Besides, it helps them to identify what students have and what they lack. Hence, developing a good classroom test is the essential activity of teachers (Atkins, Hailom, & Nuru, 1996).

On the same token, the new education policy urges teachers to use various techniques of assessment (MoE, 2002). Tests are the most commonly used techniques in different grade levels starting from lower grade levels up to higher institutions. Hence, Abatihun (2013) strongly emphasizes that “--- having the knowledge and skill of test development and principles of test construction is very crucial which is expected from all teachers or instructors” (p.301). This indicates that teachers are responsible to develop and administer qualified tests to assess students’ performance successfully. This means they should take into account the importance of validity, reliability, practicality, fairness and washback of a test.

On the other hand, the issue is still highly neglected by many language teachers and some test designers although testing is the fundamental part of teaching and learning and duty of teachers (Siddiek, 2010). For example, test construction is a skill that teachers should train themselves in the better way so as not to harm their students. However, it is a practical problem teachers frequently face. This may be due to lack of knowledge of testing techniques and lack of awareness of the importance of testing (ibid.).

1.2 Statement of the Problem

Teaching and testing are two faces of a coin (Desheng & Varghese, 2013). This means effective teaching is followed by effective testing and vice versa. As tests are the essential part of the curriculum, teachers should take due attention when constructing and administering tests. It is a duty of teachers to know what, why, how, when, and where to test. To begin with, teachers should design a test in a way that measures the relevant knowledge of students. The items should also be short, unambiguous, precise, and easy to understand for learners and other stakeholders in the curriculum. Besides, the test should cover the content taught by planning a table of specification (Brown, 2004; Mehrens & Lehmann, 1991). In general, teachers should never forget that tests have a powerful effect on students’ motivation. Students are greatly encouraged by tests constructed considering the criteria of a good test. Conversely, teaching brings negative washback effect if a test fails to mirror classroom teaching and lacks a good quality (Harmer, 2007).

In addition, tests should be administered appropriately. That's why Ur (1996) emphasizes that "How the test is actually administered and returned can make a huge difference to motivation and performance, in particular, sensitive presentation of a test can significantly reduce learner anxiety" (p.42). It shouldn't be administered in a noisy environment and disorganized classroom where there is no enough light, desks, blackboards, and etc.

On the contrary, teacher-made English language tests constructed in SCPs of Jeju Woreda, where this research was carried out, looked as if they lack good quality from the experience of the researcher. This was because of these tests, most of the time, contained items that are vague, unreadable, ambiguous, invalid, unreliable, ungrammatical, misspelled, etc. They looked they were designed carelessly without considering the principles of test item construction. In addition, teachers seemed to lack knowledge of test administration procedures and the tests were not administered in a conducive environment. Therefore, educational stakeholders complained about the tests that were being prepared and administered by English language teachers.

There are various researches which were conducted on the issues of language test items construction and administration outside and inside Ethiopia. To begin with, Ch'ng and Rethinasamy (2013) reported that English teachers were reliant on commercial reference books in constructing exam questions and sample answers in Kuching, Malaysia. Moreover, Kolawole (2014) found out that language teacher-made tests lacked validity and reliability. Conversely, Morayo and Ohia, I.N (2014), in their study, concluded that language teachers use accepted procedure well while planning, constructing, administering and scoring their tests even though they proved that the teachers in the study area have little knowledge of test planning and construction.

In Ethiopian context, Meseret (2011) found out that teacher made tests had poor qualities and teachers were not in positions to make post-test analysis. Abatihun (2013) observed that there was a significant difference in applying test construction principles, writing good multiple choice and short answer items as a function of training background and writing good short answer items as a function of teaching experience as well. Ashenafi (2014), in his study, also discovered that most of the teachers lack the awareness of test construction guidelines; lack the skill to apply them practically and even had confusion. On the other hand, Teshome (2015) reported test items

were not prepared on the basis of test blue prints, the general principles of evaluation, and specific rules of test item construction in the study area.

Hence, teacher-made tests were not effectively prepared and administered. These researchers forwarded possible solutions to alleviate the drawbacks of test construction and administration on their part.

The aforementioned studies are different from this study in their objectives, setting, sample selection, instrument usage, method of data analysis and grade level. For instance, most of them typically focused on test construction; they ignored the influence of test administration on the quality of tests. This is because properly constructed tests may bring unreliable and invalid result if they are not administered carefully (Hughes, 2003). Besides, none of these studies were conducted in primary schools on English language tests and in the setting where this research was carried out. Moreover, the researcher observed similar problems of testing from his teaching experience as most teacher-made tests lacked the qualities of test construction and administration. These reasons initiated and motivated him to conduct the research on the practice of English language test items construction and administration in SCPSs of Jeju Woreda.

1.3 Objectives of the Study

1.3.1 General Objective of the Study

The general objective of the study was to analyze the practices of English language test items construction and administration.

1.3.2 Specific Objectives of the Study

The following were the specific objectives of this study.

1. To assess the extent at which teachers apply the principles of test items construction.
2. To identify the shortcomings of teacher-made English language test items.
3. To check out the effectiveness of administering the constructed tests.

1.4 Research Questions

This study answered the following questions.

1. To what extent do teachers apply the principles of test items construction?
2. What are the shortcomings of teacher-made English language test items?
3. How are the constructed tests administered in the study area?

1.5 Significance of the Study

The results of this study will help students so that they can get qualified test items with proper test administration procedures. Besides, English language teachers and school leaders will also be benefited to create awareness on the implementation of principles of test items construction and administration and avoid negative washback effect in the teaching and learning process. It will also assist colleges of teachers training to understand the practical issues of language testing and to work on them to avoid problems related to the construction and administration of English test items. Furthermore, it will also help other researchers as a source of information for further investigation.

1.6 Scope of the Study

The study was carried out in Jeju Woreda, which is located about 182 kilometers away from the capital city Addis Ababa, Eastern Arsi Zone of Oromia Regional State in the academic year 2016-17. The study was delimited to the construction and administration of teacher-made English language achievement tests in the government schools. Hence, it was delimited to grade 7 English language teachers, students, language department heads, English language test items and English classes in the SCPSs in the Woreda.

1.7 Limitations of the Study

It would have been more comprehensive if the study had been conducted at Zone level. However, it was confined to a single Woreda due to time constraint. Additionally, the study was also limited to the construction and administration of teacher-made English language tests in the

SCPSs of government schools. It would have been better if the research had taken into account the issue in the same level of private schools.

1.8 Operational Definitions of Terms

Administration- the process of delivering tests to students in a school.

Construction- preparation of test items.

Content- the subject matter that is taught and tested.

Item- a single question in a test.

Stream- a field of study which encompasses Afan Oromo, Amharic and English as a major field in teachers training college.

Test- a collection of items like a final examination.

Woreda- the basic administrative place in Ethiopia.

Zone- one of the administrative areas next to Woreda in Ethiopia.

1.9 Organization of the Study

This paper consists of five chapters. The first chapter presents the introduction part of the thesis which comprises background of the study, statement of the problem, objectives, research questions, significance, scope, limitation of the study, operational definitions of key terms and organization of the thesis. The second chapter deals with the review of related literature which provides the reader with information related to English language test items construction and administration principles and methodologies. The third chapter presents research design and methodology. The research design, setting, population and sampling, data collecting instruments, pilot study, data collection procedures, data analysis and interpretation procedures and ethical considerations of the study are discussed in the chapter. The fourth chapter comes with the presentation, analysis and interpretation of data collected through questionnaires, document analysis, classroom observations and interview. Finally, the summary of the findings, general conclusions and recommendations are summed up in the fifth and last chapter.

Chapter Two: Review of Related Literature

This chapter attempts to review available related literature on test item construction and administration principles and methodologies. Important points that are relevant to the research topic are also discussed as follows.

2.1 What is a Test?

A test is one of educational measurement tools in teaching. It is an activity that helps teachers recognize how well students know or can do something (Ur, 1996). Similarly, Brown (2004) defines it as a system of measuring a student's ability, understanding or performance in a given domain. A test constitutes different types of items. An item is a basic component of a test. A collection of items make up a test.

2.2 Perspectives of Testing

Testing can be seen from the perspectives of traditional testing and use-oriented testing. The main focus of traditional testing is designing qualified tests that help teachers accurately measure the knowledge of students. It considers testing as a professional field following strict rules and applications of testing. Besides, it relies on the objective type of items most of the time. In short, tests are the isolated process of testing. On the other hand, use-oriented testing targets on the issues related to the reasons for giving tests and the effects that tests have on students, education and society. Issues related to the students, teachers, methods and materials designed for the test are taken into consideration. For example, the effects of test results on students, parents and society are very essential because tests are supposed to have positive washback (Shohamy as cited by Yildirim, 2010).

In my point of view, I argue for the use-oriented view of testing because developing a qualified test is not the only process of testing. The reliability, practicality, administration and washback issues of the test can affect the quality of the appropriately designed test. Thus, constructing a qualified test is not the only means to the end of testing.

2.3 Why a Test?

The primary goal of a test is to discover how much students have achieved in the course objectives of a given teaching and learning program. It also enables teachers to increase effectiveness in making modification in their teaching methods; it gives information about what students know and to decide what to teach next. In addition, it helps to discover students' difficulty of learning and provide appropriate assistance. Moreover, a test motivates students to study and show their ability to perform certain task and provides them with a sense of achievement and progress in their learning. Furthermore, it helps to evaluate the effectiveness of the syllabus, the methods and materials the teacher is using. Besides, a test provides information for decision-making. For example, a test can be given to students to decide whether or not they should enter a college or university. Lastly, it helps teachers keep a class quiet or get learners concentrate on their learning (Bachman, 1990; Clay 2001; Fulcher, 2010; Heaton, 1990; Mehrens & Lehmann, 1991; Ur, 1996).

2.4 Types of Tests

Tests can be classified in different ways based on various criteria. According to Backman (1990) language tests can be differentiated depending on five distinctive features: purpose, content, interpretation of scores, marking procedures, and testing methods. Based up on the purpose the tests that are designed can be classified as readiness, placement, diagnosis, and progress tests. Tests can also be distinguished on the basis of the content upon which they are based on as achievement, proficiency, and aptitude tests.

Moreover, tests can be criterion or norm-referenced tests based on the frame of reference for interpreting test results. Furthermore, tests are classified as objective or subjective tests on the basis of scoring procedures. Finally, tests can be identified on the basis of specific testing methods as multiple-choice, completion, essay, dictation, cloze, etc. Hughes (2003), in addition to Backman (1990), classifies tests as direct or indirect based on approaches of test construction and as discrete or integrative tests based up on language elements they test.

2.5 Teaching and Testing

Testing is the essential part of every teaching and learning experience of a successful curriculum (Brown, 2004; Madsen, 1983). Clay (2001) also explains that the general goal of testing is to encourage learning. It also reinforces learning and motivates students to concentrate on their study. Moreover, it helps teachers to examine their teaching methods and the language they are teaching. Teaching, on the other hand, determines what to test and how to test. Hence, Heaton (1990) assures as “Both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being concerned with the other” (p.5). Generally, it is possible to summarize that testing and teaching are the two faces of a coin (Desheng & Varghese, 2013).

2.6 The Focus of Language Tests

Brown and Hudson (1998) propose that “Language testing differs from testing in other content areas because language teachers have more choices to make” (p.653). Hence, language tests focus on language skills and sub-skills such as speaking, writing, listening, reading, grammar and vocabulary. Since the current language teaching method favors the communicative approach of teaching, it is valuable for test writers to concentrate on the ability to use language for real-life communication. For example, in speaking, it is possible to test the ability to understand and respond appropriately to polite requests, advice, instructions, etc. In writing, items requiring students to write letters, memos, reports and messages can be designed. In listening and reading tests, whereas, questions in which students show their ability to extract specific information can be constructed. Vocabulary tests are concerned with word meaning, word formation and collocation. Grammar tests, on the other hand, measure students’ ability to recognize appropriate grammatical forms and to manipulate structures (Heaton, 1990).

2.7 Qualities of a Good Test

Brown (2004) explains that “Constructing a good test is a complex task involving both science and art” (p.4). Heaton (1990), on the other hand, warns “A good test should never be constructed in such away to trap the students into giving an incorrect answer” (p.14). Hence, Classroom teachers are required to construct qualified tests so as to be fair and bring positive washback in

the teaching and learning program. The following are different types of qualities of a test suggested by different scholars.

2.7.1 Validity

Validity is the extent to which a test measures what it is intended to measure. Fives and DiDonato-Barnes (2013) emphasize that “Teachers need to be concerned that the test measures an adequate sampling of the class content at the cognitive level that the material was taught” when developing a test (p.1). This is because “The cornerstone of classroom assessment practices is the validity of the judgments about students’ learning and knowledge” (Wolming & Wilkstrom as cited by Fives & DiDonato-Barnes, 2013:6). Brown (2004) also assures that validity is “--- the most significant cardinal principle of assessment and evaluation” (p.30).

In addition, Siddiek (2010) warns that “A test with no content validity would spoil the whole educational process” (p.142). However, Heaton (1990:) complains that teachers “--- construct test items which succeed only in trapping students by including grammatical and vocabulary items which have never been taught”(p.14). For instance, Albussairi (2009), as cited by Siddiek (2010), reported that students did not read their textbooks and they did not benefit from their English programmes in the study area because the test did not frame the learners' behavior towards the syllabus. Hence, tests should be constructed based on what students know or learned in the classroom.

For classroom assessments, two sources of validity evidence are essential: evidence based on test content and evidence based on response process. Evidence based on test content, on one hand, underscores the degree to which a test or any assessment task measures what it is designed to measure (Wolming & Wilkstrom as cited by Fives & DiDonato-Barnes, 2013). Thus, classroom tests must be associated with the content (subject matter) taught in order for any of the judgments made about students’ understanding and learning to be concrete.

Response process evidence, on the other hand, is concerned with the alignment of the kinds of cognitive learning outcomes (knowledge or recall, comprehension, or understanding application, analysis, synthesis and evaluation) required of students during instruction and during assessment activities. The response process ensures that classroom tests assess the level of cognitive learning

outcomes that were required for students during their instructional experiences. To achieve the evidence of response process, Mehrens and Lehmann (1991) suggest that the teacher should develop qualified instructional objectives that are related to course content and eventually to the kinds of test questions the teacher proposes to use for measuring the degree of student mastery of these objectives.

In addition, it is important to identify and frame classroom objectives appropriately and they should be represented in the form of test specifications (Brown, 2004). Clay (2001) also assures that “The matching of learning objective expectations with certain item types provides a high degree of test validity: testing what is supposed to be tested” (p.5).

Another mechanism by which teachers achieve validity is by developing a table of specification. Fives and DiDonato-Barnes (2013) propose it “--- provides a strategy for teachers to improve the validity of the judgments they make about their students from test responses by providing content and response process evidence”(p.4). Hence, it can be used in conjunction with lesson and unit planning to help teachers make apparent connections between planning, instruction, and assessment. Brown (1996), on the other hand, recommends that teachers should also be very careful about planning the test objectives and specifying the types and proportions of items that will appear.

2.7.2 Reliability

Richards and Schmidt (2002) define reliability as “--- a measure of the degree to which a test gives consistent results” (p. 454). A test is said to be reliable when similar results are scored after a well-prepared test is given to similar students at different occasion. A number of factors may contribute to the unreliability of a test. These are problems in the student, in scoring, in test administration, and in the test itself (Mousavi as cited by Brown, 2004). In the construction of the test, long tests, for example, make students fatigued. Poorly written test items that are ambiguous or that have more than one correct answer, items that are easily guessed without proper studying, and unclear and ambiguous instructions are some sources of unreliability (Brown, 2004; Clay, 2001).

2.7.3 Authenticity

Bachman and Palmer (1996) define authenticity as "The degree of correspondence of the characteristics of a given language test task to the features of a target language task" (p. 23). Many test item types fail to simulate real-world tasks. They may be contrived or artificial in their attempt to target a grammatical form or a lexical item. Writing sequences of items that are not related to one another lack authenticity.

Brown (2004) recommends "The test task should be enacted in the real world, not artificial language" (p.28). For instance, reading passages should be selected from real-world sources that test-takers are likely to have encountered or will encounter. It is possible to achieve authenticity in the following ways as he has suggested.

- The language in the test is as natural as possible;
- Items are contextualized rather than isolated;
- Topics are meaningful (relevant, interesting) for the learner;
- Some thematic organization to items is provided, such as through a story line or episode; and
- Tasks represent, or closely approximate, real-world tasks.

2.7.4 Practicality

An effective language test is practical. This means the test should be readily functional by teachers with limited time and resources (Broughton, Brumfit, Flavell, Hill, & Pincas, 1980). Expensive and time consuming tests are impractical. On the other hand, the value and quality of a test sometimes can be affected by practical considerations (Brown, 2004). Hence, a practical examination is one that does not place an unreasonable demand on available resources (CELA, 2016).

2.7.5 Washback

Hughes (2003) defines washback as "--- the effect of testing on teaching and learning" (p. 1). A test should offer useful washback to the learner (Brown, 2004). This is because it has important effects and consequences within the educational system and within society more widely. Students, in particular, are affected because the results of tests are used to make important decisions about them which can affect their lives (CELA, 2016).

2.7.6 Fairness

Fairness refers to the extent to which a test measures students' ability in a way that does not give an advantage to factors irrelevant to the teaching and learning process in schools (Clay, 2001). A test should be fair to those who rely on the results of a test by performing consistently and by accurately assessing the ability being tested (CELA, 2016). This is because "Unspecified expectations and verbal tricks are unethical in testing situation" (ICDR, 2002:214).

Clay (2001) recommends the following as the basic rules of fairness: Test questions should reflect the language that is used in the classroom; Test items should be stated in simple, clear language, free of nonfunctional material and extraneous clues; Test items should also be free of race, ethnic, and sex bias; Test questions should reflect the objectives of the unit; Tests should allow teachers to understand what needs to be taught next; Expectations should be clearly known by the students; Each test item should present a clearly formulated task; one item should not aide in answering another; ample time for test completion should be allowed; and conducive environment should be created before the test is administered.

Generally, the process of the construction and administration of a good test should take into account the diversity of students and the need to treat them all fairly irrespective of their background. Test materials should also be kept securely. Unless test questions are kept securely, "--- their answers will be based on memory rather than language ability" (CELA, 2016:7). Then, the validity and reliability of test will be undermined.

2.8 Essentials in Test Item Construction

Constructing a good test is not a matter of chance. To develop efficient test items, it is very important to know the value of learning objectives, table of specification, and test item formats. They are discussed as follows:

2.8.1 The Role of Learning Objectives in Test Item Construction

ICDR (2002) defines learning objectives as statements that depict the behavior that students are expected to perform after teaching and learning process. They are included in teachers' lesson plan as statements that describe behaviors of the performance of students to be learned through lessons. They are used to identify the intended learning outcomes and help to select appropriate

instructional methods, materials, and activities. They also help to communicate parents, students, and administrators or others what is expected of the students. Besides, they help teachers to plan the construction of good test items.

Mehrens and Lehmann (1991) emphasize “Without objectives the teacher is at a loss to know both what is to be taught and hence what is to be measured” (p. 58). One of the major deficiencies in teacher-made tests results from inadequate attention being paid to the expression of instructional objectives in terms of student behavior. Therefore, teachers should state assessable, overt and unambiguous learning objectives and categorize them in the cognitive domains of Bloom’s taxonomy so as to construct valid test items because the cognitive domain categories are the main aspect to be assessed than the affective and psychomotor domain in achievement tests (ibid). The taxonomy of educational objectives for the cognitive domain helps categorize objectives at six levels of cognitive complexity: knowledge, comprehension, application, analysis, synthesis, and evaluation (Borich & Kubiszyn, 2013).

Hambleton as cited by Fulcher (2010), also explains learning objectives should “--- describe vividly the knowledge and skills to be tested and provides the framework needed to write valid test items and to evaluate item–objective congruence” which refers to the relationship between the item or task and the learning objective that it is designed to test (p.134). This matching provides a high degree of test validity (Borich & Kubiszyn, 2013; Clay, 2001). Therefore, Brown (2004) recommends that “The primary task of designing a test [---] is to determine appropriate objectives” (p.49).

2.8.2 The Role of Table of Specification

A table of specification (Henceforth, ToS), sometimes called a test blueprint, test grid, or content-validity chart, is a table that helps teachers relate their instructional objectives, the cognitive level of instruction, and the amount of the test that should assess each objective (Nortar et al. as cited by Fives & DiDonato-Barnes, 2013). It is also a tool that helps teachers “--- map the amount of class time spent on each objective with the cognitive level at which each objective was taught thereby helping teachers to identify the types of items they need to include on their tests” (Fives & DiDonato-Barnes, 2013:4).

Borich and Kubiszyn (2013), on one hand, argue that “The test blueprint conveys to the teacher the number of items to be constructed as per objectives, their level of cognitive complexity in the taxonomy, and whether the test represents a balanced picture based on what was taught” (p. 124). According to Mehrens and Lehmann (1991), on the other hand, “The purpose of the ToS is to define as clearly as possible the scope and emphasis of the test, to relate the objectives to the content, and to construct a balanced test” (p.63). The importance of a ToS shows that it is the indispensable precondition to writing efficient test items. Similarly, Alade and Omoruyi (2014) also reported a ToS is relevant to educational assessment in the study area.

Even though assembling test blueprints requires some time and effort, the ToS should be prepared before the beginning of a lesson or unit for most of its benefit (Borich & Kubiszyn, 2013; Mehrens & Lehmann, 1991). Additionally, it is also possible to draw up a table of specification before starting to write any test items (Heaton, 1990).

Therefore, Heaton (1990) recommends that the teachers should sketch detailed ToS showing features of the skills being tested and giving a comprehensive coverage of the specific language elements to be included. Backman (1990) adds that “The more detailed the ToS, the easier it should be to assure such equivalences in content” clearly (p.154).

2.8.3 Test Item Formats

In language testing, Brown (1996) categorizes test item formats as receptive and productive response items. He explains receptive response items are used to test receptive skills: reading and listening, while, productive response items are used to test productive skills such as speaking and writing. True-false, multiple choice, and matching items are included in response item formats, whereas, fill-in, short-response and task types are incorporated into productive response items. Brown and Hudson (1998) also renamed the category of test items as selected and constructed response items respectively as Brown (1996) mentions. Productive (constructed) response items encourage students to create language by writing, speaking, or doing something else but receptive (selected) response items do not. The former are easy to construct but very difficult to mark when they are marked appropriately; however, the latter require more careful preparation and consume much time although scoring them is relatively fast, easy, and objective. On the

scale of guessing, receptive response items are more susceptible than the productive ones (Brown, 2004; Brown, 1996; Brown & Hudson, 1998).

Mehrens and Lehmann (1991), on the other hand, classify test item formats as objective and subjective test items based on scoring. They group true-false, multiple choice, matching items and short answer response items under objective test items, whereas, essay items are included under subjective test items. They also propose factors to be considered in deciding which item format to use include the following: the purpose of the test, the time available to prepare and score the test, the number of pupils to be tested, skill tested, difficulty desired, the physical facilities available for reproducing the test, the age of the pupils, and the teachers' skill in writing the different types of items.

Generally, the basic principles of test item formats should be considered when tests are developed. Hence, the principles of the most common test item formats such as multiple choice, true-false, fill-in or completion, and essay items are discussed as follows.

2.8.3.1 Multiple Choice Items

Multiple choice items are the most popular items in testing. They are made up of a stem, which provides the question or problem and followed by three to five alternatives. One of these alternatives is the correct answer and the others are distractors or incorrect response (Brown, 1996; Clay, 2001; Fulcher, 2010). Heaton (1990) proposes as they test knowledge of grammatical forms and vocabulary than the ability to use language. These test items are also useful in testing higher level of cognitive domains (application, synthesis, analysis, and evaluation) when they are constructed carefully (Clay, 2001). There are various types of multiple choice items: one correct answer, the best answer, analogy type, reverse type and alternate-choice as suggested by Mehrens and Lehmann (1991).

Multiple choice items are very effective; versatile at all levels, reduce guessing, afford excellent content sampling, offer valuable diagnostic information, and provide greater test reliability per item. Nevertheless, they encompass serious limitations. For instance, they encourage cheating. They are also difficult to construct and consume time because is not easy to get plausible distractors. Unless the distractors are reasonable, the items become a good source of guessing

and superfluous (Clay, 2001; Mehrens & Lehmann, 1991). Heaton (1990) suggests "Plausible distracters are best based on mistakes in the students' own written work, their answers in previous tests, the teachers experience and a contrastive analysis between the native and target languages" (p. 32).

In addition to all of the above, scholars such as Borich & Kubiszyn (2013), Brown (2004), Brown (1996), Brown & Hudson (1998), Clay (2001), Heaton (1990), Mehrens & Lehmann (1991) and Zimmaro (2004) have forwarded guiding principles on how to construct efficient multiple-choice items. What should be done is summarized as follows.

- Multiple choice items should be constructed using context because it provides students with enough background knowledge and details to avoid ambiguity and alternative interpretations. Decontextualised and isolated sentences in a multiple-choice test simply add to the artificiality of the test situation and give rise to ambiguity and confusion. Hence, both linguistic context and situational context are essential. Moreover, an authentic language should be used.
- The stem should be stated as a direct question rather than as an incomplete statement. There should be a definite, explicit and singular question or problem in the stem.
- Negatively stated stems should be used sparingly. When they are used, underline and/or capitalize the negative word.
- The correct response should be randomly distributed throughout the test by including approximately equal proportion of alternatives A, B, C, and D as the correct response and patterns that can help students to guess should be avoided.
- Unintentional clues (grammatical, phonological, morphological, and so forth) that help students to answer an item without having the knowledge or skill being tested should be avoided completely. To do this, teachers should write multiple-choice items so that they clearly test only one concept or skill at a time.
- There should be certainly one correct answer, all the distractors should be plausible and all alternatives in an item should approximately be the same in length or in a similar format (i.e., all phrases, all sentences, etc.). Besides, the alternatives none of the above and all of the above should be used sparingly and capital letters (A, B, C, D) as response signs should be applied in order to avoid confusion.

- Alternatives should be listed on separate lines vertically (rather than including the options as part of the stem) and on the same page so that all options can be clearly distinguished.
- Each item should be designed to measure a specific objective. In order to avoid needless redundancy, it is better to check again and again.

2.8.3.2 True-False Items

True-false items are items typically written as statements and students must decide whether the statements are true or false. As of Heaton (1990), the true/false test items are widely used to test reading comprehension at elementary levels. In addition to this, Ur (1996) suggests as they are used to test listening comprehension. The items are used to test lower level of cognitive domain: knowledge (recall) and comprehension (understanding).

True-false items are advantageous in testing large amounts of content. When compared to other selected response items; they can also be constructed, administered and scored easily and quickly. For example, students can answer 3-4 questions per minute. Besides, they are good for young students. However, they encourage guessing and cheating and fail to discriminate students. They are more susceptible to ambiguity and misinterpretation as well than any other selection-type objective item (Borich & Kubiszyn, 2013; Brown, 1996; Heaton, 1990; Brown & Hudson, 1998; Clay, 2001; Mehrens & Lehmann, 1991; Ur, 1996).

These scholars have forwarded guiding principles on how to construct efficient true-false items. The guidelines are summarized as follows.

- True-false items should be based upon statements that are absolutely true or false.
- The item statement should be expressed in simple and clear language with a single idea in each test item. Be aware that extremely long or complicated statements will test reading skill rather than content knowledge.
- Enough background information and qualifications should be included so that the ability to respond correctly to the item does not depend on some special or uncommon knowledge.
- True and false items should be made of approximately equal average length throughout the test and randomize the sequence of true and false statements so that there is no discernible pattern of answers.

- Any ambiguities and unfamiliar vocabulary that might cause the students to miss it for the wrong reasons should be avoided while wording the statements.
- Absolute clues (modifiers) should be avoided because they allow students to answer correctly without knowing the correct response. They include terms like all, always, absolutely, never, rarely, most often, and so forth.
- Lifting statements from the text, lecture or other materials should be avoided so that memory alone will not permit a correct answer.

2.8.3.3 Matching Items

Matching items are items with two columns of information; the left-hand column which consists the problems is called a stem or premise, whereas, the right-hand column is called response or option. Thus, the students are expected to match the correct option to each premise (Brown, 1996; Clay, 2001). These items are used to test lower levels of cognitive domain (knowledge and comprehension) when they are appropriately constructed. Ur (1996) and Brown and Hudson (1998) explain as the items are usually used to test vocabulary knowledge (i.e., the students' ability to match definitions to vocabulary items).

Matching items are beneficial: they are efficient in testing knowledge; they are relatively compact in terms of space; and they have a low guessing factor. However, they are not appropriate for measuring higher levels of learning. They are also difficult to construct due to lack of homogenous contents (Brown, 1996; Brown & Hudson, 1998; Clay, 2001).

The scholars also suggest the following guidelines to construct efficient matching items:

- Make sure your teaching objectives allow matching items.
- Directions which clearly state the basis for matching the stimuli with the responses should be included and explained whether or not a response can be used more than once. There must be a correct response for one stem although a response may be used as the correct answer for more than one stem.
- Items that share the same foundation of information (i.e. the premises and options should be logically related to one central theme) should only be used because unrelated topics included in the same matching may allow for obvious matches and mismatches. For example, lining up definitions and the related vocabulary items is a good idea.

- The column of stimuli on the left should set the question clearly and grammatical or other clues should be avoided. All matching items should appear on the same page in order to avoid confusion.
- The two columns should not be equal in number to avoid unnecessary guessing strategy. Hence, more options should be supplied than premises or stems.

2.8.3.4 Short Answer Items

Short answer test items are questions that students answer in a few words, phrases or sentences. They are used to test recall rather than recognition. Completions and fill in-the blanks are another common labels for short answer items (Ebel & Fresbie, 1991). The difference between short answer and completion or fill-in items is that the former may require a sentence or a paragraph to fully answer the question, whereas, the latter usually require a word or phrase to complete the blank space. Completion or fill-in items, on the other hand, are more objective than scoring short answer response items (Clay, 2001).

Short answer test items are less affected to guessing than other objective test items. They are also useful in areas such as spelling and foreign language evaluations and are constructed easily (Mehrens & Lehmann, 1991). At the time of construction, they should be phrased as clear and direct questions and there must be only one concise answer or set of answers (Brown, 1996).

Completion or fill-in items usually test grammar or vocabulary. They are fairly easy to construct, are flexible in what they can assess, and are quick to administer. Heaton (1990), on his part, suggests carefully constructed completion items are a useful means of testing a student's ability to produce acceptable and appropriate forms of language. They are also frequently preferable to multiple-choice items since they measure production rather than recognition, testing the ability to insert the most appropriate words in selected blanks in sentences. Hence, these items should be constructed on the basis of the following guidelines forwarded by Brown (1996) and Clay (2001).

- So many words should not be omitted from the statement that the intended meaning is lost. Hence, teachers should omit only significant words from the statement.
- Obvious and grammatical clues to the correct response should be avoided.

- Much more credit should be given for completions than for true-false or matching items and there must be only one concise correct response.
- Enough contexts and explicit directions should be provided for clarity.
- The blank should be better put at the end of a statement rather than at the beginning and its length should be the same throughout the statement or paragraph to stop unnecessary guessing.

2.8.3.5 Essay Test Items

Essay test items are items that require students to produce, organize and communicate thoughts to answer questions. They are used to measure higher order of cognitive process: application, synthesis and evaluation (Clay, 2001; Ebel & Fresbie, 1991; Mehrens & Lehmann, 1991). The items test students' general ability of writing (Ur, 1996). According to Mehrens and Lehmann (1991), they are subdivided into two major types: extended and restricted- response depending on the amount of freedom given the students to organize their ideas and write answer. In the former type, no bounds are applied to students' points and the type of organization to be used, whereas, in the latter type, the students are more limited in the form and scope of their answer because they are told how to do specifically in the context.

Essay test items present a realistic task to the students: allow them to demonstrate ability to organize knowledge, express opinions and show originality. Therefore, they are not susceptible to guessing. They also offer students the opportunity to use their own judgment, writing styles, and vocabularies. Besides, they are easy to construct. Unfortunately, marking essay items is very difficult and time-consuming. Essay items also permit only a limited sampling of content (low validity).

The following are guidelines that help teachers to construct effective essay items (Clay, 2001; Ebel & Fresbie, 1991; Mehrens & Lehmann, 1991).

- The task and its instruction should be clearly defined.
- Optional questions should be avoided on an essay test because it makes it difficult to discriminate between the students who could respond correctly to all items.

- Essay items should be written at different levels of learning because the item should represent a situation that tests the student's ability to use knowledge in order to analyze, justify, explain, contrast, evaluate, and so on.
- An appropriate amount of time should be given to students for each question, as well as the desired length and format of the response.
- Students should be informed in advance about the items, the proportional value of each item in comparison to the total grade, and the value of neatness, handwriting, grammar, etc.

2.9 General Phases of Test Items Construction

Mehrens and Lehmann (1991) strictly emphasize that “Good tests do not just happen!” (p.55). They approve that one of the major errors committed by teachers when preparing classroom tests is lack of adequate planning. Good tests require adequate and extensive planning so that the instructional objectives, the teaching methods to be employed, the content, and the evaluative procedure are all related meaningfully. However, teachers leave planning a test until the last day and rush like mad to prepare something (ibid.). Similarly, Siddiek (2010) reported as “... it was evident that teachers took very little time in preparing their tests.” (p.133). Then, the test items constructed in this manner becomes ambiguous, not scorable, or too difficult. Every test should be planned and reviewed critically by teachers to minimize the deficiencies. Essential phases of test construction are discussed as follows.

a. Defining the Purpose of the Test

At first, teachers need to decide why they are testing because identifying the purpose helps teachers what to do next. Classroom achievement tests, for example, serve a variety of purposes; judging the pupils' mastery of skills, ranking students, diagnosing students' difficulties, encouraging good study habits, and motivating students in their study (Mehrens & Lehmann, 1991).

b. Classifying the Content to Be Tested

After identifying the purpose of a test, the second major step teachers do should be to decide what knowledge and skills are going to be measured depending on the instructional objectives and what has been taught in class because what to test helps determine how to test. At this stage, teachers must consider the relationships among learning objectives, teaching, and testing (Borich & Kubiszyn, 2013; Mehrens & Lehmann, 1991).

c. Listing Learning Objectives

Appropriately designed learning objectives help teachers construct valid test items. Hence, learning objectives should be unambiguous, precise and measurable. It is important to list them down in the levels of cognitive domain (Brown, 2004; Mehrens & Lehmann, 1991).

d. Preparing a Table of Specification

After the course content and learning objectives have been specified, teachers should develop a ToS because it plays a significant role to avoid complaints that students make and improve the content validity of teacher-made tests. However, it is advisable to prepare a ToS before the beginning of teaching if possible (Heaton, 1990; ICDR, 2002; Mehrens & Lehmann, 1991).

e. Determining the Weight

Classroom teachers know the relative emphasis placed upon the various instructional objectives in the lessons. Hence, they are responsible to assign various weights although there is no clear rule that help teachers to determine the weights. The weights assigned should reflect the relative emphasis used by the teachers when they taught the course (Mehrens & Lehmann, 1991).

f. Relating the Test Items to the Instructional Objectives

Borich and Kubiszyn (2013) strongly recommend that “The learning outcome and conditions specified in the test question must match the learning outcome and conditions described in the objective” (p.114). It ensures that the test teachers develop will have content validity. To acquire this goal, “There must be a perfect match” as Mehrens and Lehmann (1991:64) advise.

g. Selecting Appropriate Test Item Format

Once the instructional objectives have been sorted out, the next step is to select appropriate item types to measure them. The combination of essay and objective items (receptive and productive response items) are recommended to devise because different skills and abilities are measured by each of them (Borich & Kubiszyn, 2013; Brown, 1996; Mehrens & Lehmann, 1991).

h. Formulating Well-defined Test Tasks

Test items constructed without proper plan are awkward. They should not be vague and ambiguous. They should be grammatically correct and free from spelling and typing errors. In addition, they should be free from textbook or stereotyped language, excessive wording, complex sentences, double negatives, inappropriate and biased vocabulary. Moreover, test items should be contextually devised and free from specific determiners. Besides, a test item should have only one correct answer, items should be arranged from easy to difficult and all parts of the item should be on the same page. If negatives must be tested, wise test writers emphasize the negative elements (by underlining them, typing them in capital letters, or putting them in boldfaced type) so that students are sure to notice what is being tested. Students should not miss an item because they did not notice a negative marker (Borich & Kubiszyn, 2013; Brown, 1996; Mehrens & Lehmann, 1991).

i. Editing Test Items

At this phase of testing, teachers need to proofread their test items or let other teachers comment on them because tiny mistakes can cause serious problems. Hence, they should make sure that directions are absolutely clear and concise, each item measures a specified objective, each item is stated in clear and simple language, distractors of multiple-choice item are appropriate, the difficulty of each item is appropriate for students and the language of each item is sufficiently authentic (Borich & Kubiszyn, 2013; Brown, 1996; Mehrens & Lehmann, 1991).

j. Determining the Test Length

Items should be large enough to provide an adequate sample of instructional objectives and content areas. The purpose of the test, the kinds of items used, the reliability desired, the age and

ability of the pupils tested, the time available for testing, the length and complexity of the item, the amount of computation required, and the instructional objective tested determine how many items should be used (Mehrens & Lehmann, 1991).

k. Determining the Item Difficulty

Excessively difficult items frighten students. Difficult test items, however, should be included to challenge clever students and to help them study harder. The purpose of the test, the ability and the age or grade level of the students should be considered to determine the difficulty of the items (Borich & Kubiszyn, 2013; Mehrens & Lehmann, 1991).

l. Deciding When to Test

Even though tests are not the only method of assessment, they should be given to students at least twice a semester to assess the progress of students. However, teachers can administer frequently if there are enough resources and time (Mehrens & Lehmann, 1991).

2.10 Drawbacks in Teacher-made Tests

Brown (2004) emphasizes that “Tests need not be degrading, artificial, anxiety-provoking experiences” (p.2). Hence, they should be constructed and administered following the guiding principles so as to bring beneficial washback. However, Mehrens and Lehmann (1991) conclude that “By and large, teacher-made achievement tests are quite poor. But that shouldn't be surprising or unexpected” (p.52). They also found out that teachers lack understanding of measurement. Technically their tests are simplistic and depend upon short answer, true-false, and other easily prepared items. Their multiple-choice items often have serious flaws-especially in distracters (ibid).

Specifically, Gullickson and Ellwein as cited by Abatihun (2013) also mention the shortcomings related to test construction as follows.

...teachers lack sufficient training in test development, fail to analyze tests, do not establish reliability or validity, do not use a test blueprint, weight all content equally, rarely test above the basic knowledge level, and use tests with grammatical and spelling errors.(p.301)

Different studies have also approved these problems. For example, Meseret (2011) has approved that most of the test items were not constructed based on the principle of test construction in the study area. In addition, Ashenafi (2014), in his study, discover that most of the teachers lack the awareness of test construction procedures; lack the skill to apply them practically and even had confusion. Moreover, Teshome (2015) reported test items were not prepared on the bases of a test blue print, the general principles of evaluation, and specific rules of test item construction in the study area.

In their study, Ch'ng and Rethinasamy (2013) also reported that English teachers are reliant on commercial reference books in constructing exam questions and sample answers in Kuching, Malaysia. Moreover, Kolawole (2014) found out that language teacher-made tests lacked validity and reliability. Conversely, Morayo and Ohia, I.N (2014), in their study, have concluded that language teachers use accepted procedure well while planning, constructing, administering and scoring their tests even though they proved that the teachers in the study area have little knowledge of test planning and test construction. Some other specific problems that teachers commit when constructing a test are discussed as follows.

i. Ambiguous Test Items

Ambiguity occurs when a statement or word is differently interpreted by different students. Mehrens and Lehmann (1991) assert that “Differential performance should be the result of differences in knowledge of the subject matter, not of differences in the interpretation of the item” (p.53). Classroom teachers can construct ambiguous questions when they formulate test item formats. For example, teachers should be careful when constructing true-false item because the item can be true or false based on the students’ reference.

ii. Excessive Wording

Teachers made tests are prone to excessive wordiness. Teachers think that wording a question makes the question clear which is completely wrong assumption. To avoid this problem, they should use precise and clear-cut words so that students easily understand the items (Mehrens &Lehmann, 1991).

iii. Lack of Appropriate Emphasis and Content Sampling

The objectives stressed and taught in the class were not covered appropriately by teacher-made tests. Hence, the tests are heavily loaded with items that only test the student's ability to recall specific facts and information. This is because it is so much easier to prepare items that measure factual recall than it is to write test items that measure comprehension, synthesis, and evaluation. As a rule, teachers should construct items that measure an adequate sampling of the class content at lower and higher order of cognitive process (Borich & Kubiszyn, 2013; Mehrens & Lehmann, 1991).

The mismatch, on the other hand, leads to a test that fails to provide evidence from which teachers can make valid judgments about students' progress (Brookhart as cited by Fives & DiDonato-Barnes, 2013). Therefore, classroom tests must be aligned with the content (subject matter) taught and reflect learning objectives in order for any of judgments that teachers make about student understanding and learning to be meaningful. The best strategy teachers can use to alleviate these problems is to develop a ToS because it can aid immensely the construction of valid and balanced test items (Fives & DiDonato-Barnes, 2013).

However, Mekonnen, Julia & Deepika (2014) reported as ToS are not prepared by all of the school teachers before test construction and teachers simply prepare tests or exams by looking at the major contents of the textbooks ignoring the weight given to each content and such practice of the school teachers signify their test lacks validity. Alade and Omoruyi (2014), on the other hand, reported that there is lack of coherence in the preparation of ToS which encourages content validity.

iv. Use of Inappropriate Test Item Formats

Test item formats are not selected arbitrarily. Firstly, they should be matched with the learning objectives. Then, appropriate items are selected considering the peculiar characteristics of each test item. However, teachers use inappropriate test item formats most of the time. Teachers should decide which type of item is used to assess each objective at the correct level. They should also take into account the fundamental considerations in selecting test item formats. On the other hand, a one-item multiple-choice test would never seem fair or reliable to any teacher

or student because single-item test could never do a really good job of testing (Brown & Hudson, 1998).

v. Decontextualized Test Items

The sequence of items in the decontextualized tasks takes students into different topic areas with no context for any. Language teachers, especially, are expected to construct items with in a context because the sequence of items in the contextualized tasks achieves some degree of authenticity by contextualizing all the items in a story line. For example, conversations can be one input for contextualized test items (Brown, 2004).

vi. Inauthentic Language

One of the problems teachers face in language testing is unable to use authentic language. Their test items fail to create real-world tasks. They may be contrived or artificial in their attempt to target a grammatical form or a lexical item. The sequence of items without any relationship lacks authenticity. Therefore, teachers should construct test items that mirror language use in real life (Brown, 2004).

2.11 Test Administration

Test administration is the overall test procedures that teachers apply before, during and after the test. What an administration of a test includes is described by Fulcher (2010) as follows.

It includes everything from having enough copies of the test paper and answer sheets, ensuring that not a single copy is 'misplaced', and transporting the materials to a test venue that has been cleaned the evening before and set out according to written regulations so that desks are a set distance apart. Then there is arranging the correct number of invigilators/proctors for the number of test takers expected, ensuring that doors are unlocked, equipment (including lights, heating, air conditioning) is working properly, and that candidate identity can be checked and registered against a number that allows the matching of scores (following blind marking) to individuals.(p.253)

This indicates that test administration is not a simple process. It needs a due attention because the best test may bring unreliable and invalid result if a test is not administered carefully (Hughes, 2003). Fulcher and Davidson (2007) also suggest “Any mistakes, inconsistencies or

abnormalities at any stage in the test administration process can threaten validity” (p.115). Hence, the administration of any test needs careful consideration and planning.

Test administration gives the suggestions and preparations to help students psychologically and academically prepare for the test and attend all the tests. The usefulness and interpretability of test scores require a test to be administered and scored according to the developer’s instructions (Fulcher, 2010).

Test Administration is the easiest of all testing process if and only if it is free from problems encountered (Ebel & Fresbie, 1991). This is because administration conditions have an adverse effect on students’ performance (Cohen & Wollack as cited by Fulcher, 2010). Hence, there are problems that may arise from the test preparation, psychological aspects, physical and environmental conditions, time and place of testing, and other extraneous variables.

2.11.1 Test Preparation Considerations

Teachers, at first, should think of its practicality when they are constructing test items because practicality is one of the qualities of a good test. For example, a too long test is unwanted because it brings administration difficulties like mental strain and tension among the students taking the test (Heaton, 1990).

Each test item should also be constructed clearly following the principle of writing a test item and test item format. Then, the items should be grouped into similar formats and arranged from easy to difficult. This enables students to answer items correctly by building confidence and reducing test anxiety (Borich & Kubiszyn, 2013; Mehrens & Lehmann, 1991). With regard to the sequence of test item formats, Gronlund, as cited by Mehrens and Lehmann (1991), suggests the following order based on the complexity of instructional objectives they measure: true-false, matching, short answer, multiple choice and essay items.

On the other hand, the items should be spaced properly and their stem and options (in case of matching and multiple choice items) should be on the same page. If there are supporting materials like diagrams, maps, or other, they should be positioned immediately above the item or items to which they refer. The answer keys should also have fairly random patterns. They should

be distributed about equally between true and false and among multiple-choice options as well (Borich & Kubiszyn, 2013; Mehrens & Lehmann, 1991).

Test directions, in addition, should be vividly and concisely written. They should include the basis on which to select the answer, how to record answer, the time to be allotted for various sections and the value of the items. Heaton (1990) stresses “Unless all students are able to follow the instructions, the test will be neither reliable nor valid” (p.168).

Finally, the test should be proofread, for example, the typographical and grammatical errors, before they are going to be reproduced. It is advisable to review it independently or let other teachers who teach the same subject proof read it. Having to announce corrections to the class just before the test or during the test will waste time and confuse students (Borich & Kubiszyn, 2013; Mehrens & Lehmann, 1991).

At the stage of reproduction, the copies reproduced must be neat, tidy and aesthetically pleasing because they affect the validity and reliability of the test. Hence, nothing is worse and more bewildering to the students than an untidy test paper, full of misspellings, omissions and corrections (Heaton, 1990). This, surely, helps students build confidence to read the items and answer correctly. It also helps them to use the given time appropriately.

2.11.2 Psychological Aspects

To maximize the reliability of a test, teachers need to create a positive mental attitude in the students that are going to take the test because the students cannot do their best when they are tense and nervous. Their performance is influenced by mental attitude and motivation (Mehrens & Lehmann, 1991). Ur (1996) emphasizes as tests should be announced at least a week in advance in order to give students plenty of time to prepare and they should be told when, where and how long the test will be. They should also be informed the content to be tested, what sort of items formats will be used, and how answers will be assessed. This increases motivation and avoids test anxiety which affects the performance of students.

2.11.3 Physical and Environmental Conditions

Physical and environmental conditions and test performances are interrelated. It is obvious that the former affects the latter. Cohen and Wollack, as cited by Fulcher (2010), accentuate the issue of these conditions as:

It is reasonable to expect that administration conditions have a non-negligible effect on examinee performance. If it is too hot or too cold in the testing room, performances of some examinees are likely to be negatively affected. Similarly, if it is too noisy, some examinees may be distracted and perform below their potential. (p.255)

Brown (2004) also recommends that street noise outside the building, photocopying variations, the amount of light in different parts of the room, variations in temperature, and even the condition of desks and chairs can be sources of unreliability of a test if not considered while administering a test. Hence, the physical and environmental conditions should be as comfortable as possible and the students should be as relaxed as possible (Mehrens & Lehmann, 1991).

2.11.4 Time and Place Considerations

Time of testing affects test performance Backman (1990). Hence, it is not advisable to give a test when the students are doing pleasant things, such as lunch, recess, gym, etc. Teachers should not also give tests immediately before or after long vacations or a holiday, or the day before or after a championship game (Mehrens & Lehmann, 1991). On the other hand, enough time should be allotted to do the test and students should be given a warning 15, 10, and 5 minutes before the time limit is up, so they are not caught at the deadline (Borich & Kubiszyn, 2013; Heaton, 1990). With regard to place, tests should be given in a familiar place or else students should be informed the place in advance if there is a change.

2.11.5 Considerations of Extraneous Variables

Extraneous variables are simply any factors that affect test scores which are not related to what the test is anticipated to measure (Fulcher, 2010). For example, cheating is the most serious problem that helps students score by deceitful means. It is teachers or proctors responsibility to control cheating (Ebel & Fresbie, 1991). Mehrens and Lehmann (1991) suggest that careful proctoring is the single most effective method to minimize, eliminate, discourage, and detected it. Nevertheless, many teachers do not take their proctoring responsibilities seriously enough.

They are physically present in class but spend their time doing something else such as reading a novel, writing a letter, grading papers or etc. It is also believed that cheating is avoided by a well-planned test which covers appropriate content sampling from lower to higher level of cognitive process. Fulcher (2010), on the other hand, recommends “--- preparing seating chairs, checking candidate identity properly, spacing seats to avoid copying, and the appropriate monitoring of students during the test” help to prevent cheating (p.264). Borich and Kubiszyn (2013), on their part, advise teachers or proctors to “--- inform students about penalties for cheating and implement the penalties when cheating occurs” (p.224).

Test security, on the other hand, is also another factor that affects test reliability and administration. Hence, it should be appropriately maintained throughout the testing process because it helps to ensure that no one has an unfair advantage (Fulcher & Davidson, 2007). They also affirm that training on how to administer a test should be given to invigilators or proctors.

Chapter Three: Research Design and Methodology

This chapter presents the research design and methodology. The research design, setting, population, samples and sampling techniques, data collecting instruments, pilot study, data collection procedures, data analysis and interpretation procedures and ethical considerations of the study are described in brief.

3.1 The Research Design

The main objective of this study was to analyze the practice of English test items construction and administration. In order to achieve this objective, a descriptive survey research design was employed. This was because it is targeted at the description of the current phenomena in education (ICDR, 2002; Kothari, 2004; Lodico, Spaulding & Voegtle, 2010; Singh, 2006).

In this study, quantitative research method supplemented with document analysis, semi-structured interview and English class observation was utilized. The purpose of using supplementary tools of qualitative methods was to validate the quantitative data which were obtained from the questionnaire. To obtain the relevant data, therefore, the researcher used questionnaire for English language teachers and grade 7 students, document analysis for teacher made English language test papers, semi-structured interview for language department heads and English class observations for test administration.

3.2 Research Setting

This research was conducted in Jeju Woreda, Arsi Zone of Oromia Regional State. This is chosen as the setting because it is the researcher's work place. Jeju Woreda is one of the twenty-five Woredas found in the zone. The Woreda is bordered by Sire Woreda in the West, Boset Woreda of East Shoa Zone in the North, Merti Woreda in the East and Guna and Sude Woredas in the South. Arboye is the capital town of the Woreda which is located about 182 kilometers away to the South of the capital city, Addis Ababa. There are twenty six rural kebeles and three towns with the population of 124,041 (Jeju Woreda Communication Office, 2016/17). Among the total of 53 schools clustered in 13 cluster resource center found in the Woreda, the study area of this research included 32 SCPSs.

3.3 Population , Samples and Sampling Technique

English language teachers, language department heads, teacher-made English language test papers, English classes and grade 7 students in SCPs in Jeju Woreda were the population of this study. The researcher used census sampling technique to select all (45) English language teachers who were teaching English in SCPs. Pertaining to this, Lodico et al. (2010) mention as it is a non-random sampling technique which allows the total population to be included in a sample when the number is not too large in quantitative research.

Teacher-made English language test papers that were analyzed include test papers which were prepared and administered in the first and second semester of the academic year 2015/16 and in the first semester of 2016/17 in corresponding CRC schools. Hence, the total population of the test papers was 39. From these, 13(33.3%) test papers in the CRC schools were taken by simple random sampling technique to give equal chance.

The test items were prepared and approved at CRC schools. Thus, 8 (61.5%) language department heads from the 13 CRC represented schools were selected using simple random sampling methods so as to conduct semi-structured interview hoping to get profound information. From 3 (23%) CRCs, 3 (10%) schools which consisted of 7 English classes were also selected using simple random sampling technique for the purpose of class observations of test administration.

There were 438 grade 7 students in the selected schools. Thus, 88 (20%) students were selected by using simple random sampling technique. The sampling size of the students and schools was determined based on the suggestion given by Singh (2006). He proposes that 10-20% of the accessible population for the sample can be used for descriptive research. The researcher used simple random sampling technique because it gives equal chance for the target population to be selected as a sample (ICDR, 2002; Kothari, 2004).

3.4 Data Collecting Instruments

The researcher employed questionnaire, document analysis, semi-structured interview and English class observations as data collecting instruments. Using multiple instruments for data gathering gives more weight to the validity of the research and helps to cross-check the data collected from different angles (ICDR, 2002).

3.4.1 Questionnaires

Questionnaires are one of the most popular data gathering tools that are used to collect data from a lot of respondents. They can also be completed at a particular time and place. Moreover, they help respondents give responses freely (Gray, 2004). Hence, questionnaires were used to gather quantitative data from English language teachers and grade 7 students in SCPSs in Jeju Woreda. 19 closed ended items (12 concerning the construction of test items and 7 with regard to test administration) were adapted and modified from Morayo and Ohia, I.N (2014) and 41 closed and 5 open ended items were prepared by the researcher based on the related literature considering the guidelines of writing questionnaire items. For example, questions that seek positive and negative responses were added to control harmful effects in data analysis (Dörnyei, 2007; Yalew, 2011).

Besides, five point rating scales containing adverbs like always (4), usually (3), sometimes (2), rarely (1) and never (0) were used to each of the closed-ended items how often the respondents utilized the principles of test items construction and administration (Dörnyei, 2007). Hence, 5 open and 60 closed ended items on the principles of test items construction and administration were prepared, piloted and distributed to English language teachers. In addition, questionnaires containing 2 open ended and 32 closed ended items on the procedures of test administration were also prepared, translated into Amharic and Afan Oromo, were piloted and distributed to grade 7 students. This tool was chosen because respondents could read and understand the questions and write down the answers by their own. For further information, see Appendix A, B, C and D.

3.4.2 Document Analysis

ICDR (2002) mentions document is a good source of information for educational research. It is easily accessible and enables the researchers to study past events and issues if it is documented appropriately. Tharenou, Donohue and Cooper (2007), in addition, recommend as “It can [---] be used for triangulation of data, helping to counteract the biases of other methods and supplement sources of information” (p.125).

Thus, 13 English language test papers which were prepared and administered in the first and second semester of the academic year 2015/16 and in the first semester of 2016/17 in corresponding CRC schools were screened with the principles of test items construction to determine the quality of the construction of the test items and for the sake of validity of the research. For this reason, the researcher prepared evaluation checklists on test items construction based on the related literature. Look at Appendix F for further information.

3.4.3 Semi-structured Interview

Semi-structured interview allows probing the views and opinions of the respondents where possible (Gray, 2004). Consequently, it was also used to collect data from language department heads in order to get in-depth information and to validate the data collected from the questionnaire, teacher-made English language test papers and class observations. Based up on this, the researcher prepared 7 questions based on the related literature and interviewed 8 (60%) English language department heads from the 13 CRC represented schools using simple random sampling method. For further information, look at Appendix G.

3.4.4 Class Observation

The researcher used semi-structured observation as a tool for gathering data concerning the procedures and considerations for test administration. This type of tool is appropriate in descriptive studies and the information obtained under the method is related to what is currently happening (Kothari, 2004). In addition, Lodico, Spaulding & Voegtle (2006) recommend as observation guidelines are essential to get concrete data. Hence, the researcher prepared guidelines based on the related literature to conduct class observations several times for English

classes test administration in order to collect data in a systematic manner. For further information, look at Appendix E.

3.5 The Pilot Study

To select the most appropriate and accurate instruments to serve for data collecting, the two criteria used to judge the quality of the pre-established instruments are reliability and validity (Lodico et al. (2006). They strictly emphasize that “If an instrument does not have sound reliability and validity the instrument is of no value” (p.88).

3.5.1 Reliability of the Questionnaires

Reliability is the consistency of scores that instruments produce for over repeated time and/or across different individuals (Lodico et al., 2010). This means it is targeted at consistency but not accuracy. To pre-test reliability of the questionnaire and foresee the possible problems that can be encountered in the administration process due to the wording of the items, consequently, a pilot test was conducted at Abomsa Number 1 and 2 SCPSs which were located outside the study area. The questionnaires were administered to 10 English language teachers and 22 grade 7 students in the aforementioned schools. Then, the researcher calculated the coefficient of Alpha (α) to check the reliability of the questionnaires. Accordingly, the researcher found out the coefficient of alpha .96 and .88 for teachers’ and students’ questionnaire respectively.

Nunnally, as cited by Tharenou et al. (2007:153), explains that coefficient alpha ranges from 0 to 1 and measures that are highly reliable have alpha coefficients of .90 or greater, while scales that have alphas below .70 can be said to have less than fair reliability. Muijs (2004) also emphasizes as coefficient of alpha measure to be over .70 of internal consistency. Generally, Lodico et al. (2006) summarize “The closer to +1.00 the reliability coefficient is, the more highly reliable the instrument” (p. 91). Therefore, the result indicates that the instruments were highly reliable.

3.5.2 Validity of Data Collecting Instruments

Validity is the degree to which an instrument measures truly what it is intended to measure (Lodico et al., 2006). It is about accuracy; not consistency. To insure the validity of the

instruments, hence, the researcher prepared them in accordance with the information in the review of related literature and research objectives. Besides, the researcher provided the questionnaire, interview questions, test items evaluation and observation checklists for the advisor and colleagues to proofread and comment on them. Moreover, the validity of the guidelines to check the English language test items construction was approved by operating them on a test administered in the piloted school. Then, the instruments were improved if there were ambiguities, jargons and unclear language which the researcher had not noticed. Questionnaire for students was also translated into their mother tongue so as to understand the items and provide their response in a better way.

At the time of piloting, teachers and students who filled the questionnaire were asked to comment on the items in the questionnaire. As a result, they forwarded some comments that need to be improved. For example, the word “vet” in item 3.7 was unfamiliar for teachers. In addition, some words were also new for students even though the questionnaire was translated into their mother tongue. The researcher valued the comments given by the respondents. Hence, the word “vet” in the teachers’ questionnaire was substituted by familiar word “examine.” Moreover, English terms were used to avoid confusion in the students’ questionnaire considering the students’ comments in the try out stage of the instrument.

3.6 Data Collection Procedures

The researcher implemented several steps to gather concrete data. Firstly, the abovementioned data gathering instruments were developed. Thus, the questionnaires and checklists for class observation and evaluation of test items construction were constructed adapting items from the report of Morayo and Ohia, I.N (2014) and the information in the related literature. Secondly, the questionnaires were administered to English language teachers who were teaching at SCPSs of neighboring Woreda to pilot the questionnaires. Thirdly, authorization letter from Adama Science and Technology University was given to Jeju Woreda Education office. Then, the researcher took the authorization letter from the office to the CRC represented schools. In the schools, negotiations were made between the researcher and the administration of the sampled schools on the issue of data collecting including the teachers. Next, sample tests prepared by the respondents were collected in order to analyze them based on the principles of test item

construction and the questionnaires were also administered to teachers in each SCPS and students in the sample schools. Semi-structured interview was also conducted with 8 language department heads as well.

The researcher also observed test administration process based on the guidelines of observation. Since tests were administered at the same time in all schools, the researcher employed 7 teacher observers by training them on how to observe the process for a day in Arboye SCPS. In addition, these teachers also evaluated the sample test papers based on the evaluation checklist prepared by the researcher. Finally, the data were gathered and prepared for the analysis stage.

3.7 Data Analysis and Interpretation Procedures

In the process of data collection and analysis, a total of 45 questionnaires were distributed to teachers who were teaching in the SCPSs of Jeju Woreda. However, 41 questionnaires were returned. Hence, the response rate was 91% for teachers' questionnaires. Conversely, all 88 questionnaires (100%) administered to students were returned. Then, the researcher screened the data gathered from the respondents before inserting them to the SPSS software. This was a very important phase of preparing data because the result of the data can be distorted completely by some mistakes, misunderstanding or carelessness of the respondents (Dörnyei, 2007; Yalew, 2011). Consequently, 2 questionnaires from teacher respondents were excluded because they were not filled appropriately and data from 39 questionnaires were analyzed in this chapter.

Furthermore, items which seek negative responses were reversely coded so as not to misrepresent the data as Yalew (2011) recommends. Such items include the 9th item in the constructing, the 1st item in the revising, and the 15th item in the pre-test administration part of teachers' questionnaire and the 15th and 20th items in the pre-test administration and the 11th item in the while test administration part of the students' questionnaire. All sample test papers (13) were collected from each CRC schools. In addition, a semi-structured interview was conducted with 8 language department heads in the CRC represented schools and 7 class observations were made several times while English language tests were being administered to grade 7 students in the sample schools.

The data gathered using quantitative and qualitative methods were analyzed both quantitatively and qualitatively respectively. The information gathered through closed-ended questionnaire was coded, tabulated and analyzed using descriptive statistics or numerical analysis such as frequency, mean and standard deviation (Henceforth SD) by using SPSS version 20 software. Besides, a percentage was used to analyze the demographic characteristics of the respondents.

In this regard, Dörnyei (2007) recommends as descriptive statistics offer a tidy way of presenting and describing data in a more professional method. He also comments as they help researchers summarize findings by describing general tendencies in the data and the overall spread of scores. Descriptive statistics, on the other hand, are used to present quantitative descriptions in a manageable and intelligible form (O’Leary, 2004).

Lewin (2005) recommends as “Mean values should always be considered alongside a measure of variability” such as variance and standard deviation (p.223). Specifically, the mean is the most common widely used descriptive measure because it takes into account all the scores, whereas, the standard deviation is an indicator of average distance of the score from the mean taking account of all of the scores and it is used mostly in research studies (Dörnyei, 2007; Kothari, 2004;Yalew, 2011). Dörnyei (2007) suggests that standard deviation score becomes high when the “...sample is heterogeneous and contains extreme score, whereas, [it] becomes low in a homogeneous sample with scores clustered around the mean” (p. 194). Frequency, on the other hand, depicts each score and the overall performance of a group (Lodico et al., 2010).

The additional data gained from the open-ended questionnaire, semi-structured interview, documents (English language test or examination papers) and English class observations were also coded, categorized and identified thematically in order to approve or disapprove the quantitative data.

3.8 Ethical Considerations

The researcher was aware of essential research ethics when conducting this research. This is because research ethics play great roles in facilitating the process of data gathering. To begin with, the data was collected with the knowledge of the concerned bodies like the Woreda education office, school principals, department heads and teachers. Moreover, the test or exam

papers were collected negotiating with them as they were only used for the purpose of the research. They were attached as appendices avoiding the names of the CRC schools and teachers who prepared them. They were also evaluated by other teachers to be free from bias. In addition, the objectives of the study were explained for the respondents in the questionnaire and interview, assurance was given that the information respondents provide to be only used for the purpose of the research and the identity of the respondents was kept confidential. Thus, the respondents weren't asked to write their name. The same thing was done for observation and for gathering data from the documents by creating friendly relationship.

Chapter Four: Presentation, Analysis and Interpretation of the Data

This chapter presents the data, analysis and interpretation of the data gathered through the questionnaire, document analysis and observation checklist. It contains seven main sections. The first section deals with the demographic characteristics of the respondents. The second discusses planning a test. The third explains the construction of a test. The fourth presents the revising stage of a test. The fifth, sixth and seventh show actions before, while and after test administration respectively.

The basic statistics that are presented in the tables below include the frequency (F), mean (M), standard deviation (SD) and the number of participants (N). Items which have similar mean and standard deviations were analyzed together, whereas, items with different score were analyzed showing the scores alone.

4.1 Demographic Characteristics of the Respondents

The respondents were inquired to indicate their personal information before providing their response to the items prepared for the study. The details are discussed as follows:

Table 1: Demographic characteristics of the respondents

№	Items		Respondents			
			English Teachers		Students	
			Number	%	Number	%
1	Sex	Male	33	84.6	46	52.3
		Female	6	15.4	42	47.7
		Total	39	100	88	100
2	Age	11-15	-	-	69	78.4
		16-20	-	-	19	21.6
		21-25	5	12.8	-	-
		26-30	11	28.2	-	-
		31-35	8	20.5	-	-
		35-40	5	12.8	-	-
		41-45	7	17.9	-	-
		≥ 45	3	7.7	-	-
		Total	39	100	88	100
3	Academic	Diploma	22	56.4	-	-
	Status	Degree	17	43.6	-	-
		Total	39	100	-	-
4	Academic	English as a major	18	46.2	-	-
	Field	English as a minor	5	12.8	-	-
		Stream	13	33.3	-	-
		English neither as a major nor a minor	3	7.7	-	-
		Total	39	100	-	-
5	Teaching	1-5	9	23.1	-	-
	experience	6-10	11	28.2	-	-
	(in years)	11-15	6	15.4	-	-
		16-20	3	7.7	-	-
		≥ 25	10	25	-	-
		Total	39	100	-	-

4.1.1 Sex

As Table 1 shows, from teacher respondents, 33 (84.6%) were males, whereas, 6 (15.4%) were females. This shows that the number of female English language teachers is small relative to male teachers. Special attention should be given to encourage females in the field. On the other hand, 46 (52.3%) and 42 (47.7%) respondents were male and female students respectively.

4.1.2 Age

As can be seen from Table 1, most of the teachers were found in the age range of 26-30, whereas, the majority of the students were in the age range of 11-15.

4.1.3 Academic Status of Teachers

Table 1 indicates that 22(56.4%) of teacher respondents were diploma holders, whereas, the other 17(43.6%) were degree holders.

4.1.4 Academic Field of Teachers

Item number 4 of Table 1 illustrates that 18(46.2%) of teachers were English major teachers and 5(12.8%) were English minor teachers. It also indicates that 13(33.3%) were stream teachers who were English, Amharic and Afan Oromo major teachers. On the other hand, English was neither major nor minor for 3(7.7%) teachers. It is possible to deduce that the majority of teachers fulfilled the necessary requirement even though there were teachers whose field was not English.

4.1.5 Teaching Experience of Teachers

Item number 5 of Table 1 shows that the majority of teachers had 6-10 followed by 21 and above 21 years of teaching experience.

4.2 Teachers' Response to the Application of Principles of Planning Classroom Tests

Table 2 shows the extent to which the principles of test items construction were implemented among teacher respondents at the planning stage of test items construction.

Table 2: Teachers' response to the extent to which they applied principles of planning classroom tests

No	Items	N	Frequency Scales					M	SD
			4	3	2	1	0		
1	I define measurable instructional objectives that help me construct qualified test items.	39	8	12	14	3	2	2.54	1.07
2	I categorize instructional objectives in the levels of domains of Bloom's taxonomy.	38	8	9	10	8	3	2.29	1.25
3	I determine the objective(s) of my test before constructing the tests.	39	15	9	9	4	2	2.79	1.21
4	I prepare tables of specification to construct tests items.	39	10	12	6	3	8	2.33	1.47
5	I prepare test drafts before constructing my tests.	39	11	13	9	3	3	2.67	1.19
6	I determine the content area to be tested before constructing the tests.	39	23	9	7	0	0	3.41	.78
Grand mean and SD								2.67	1.16

Note: N= Number of respondents who filled the items correctly.

0= Never, **1**= Rarely, **2**= Sometimes, **3**=Usually and **4**= Always

M= Mean **SD**= Standard deviation

Item 1, 3, and 5 of Table 2 indicate that English language teachers almost usually used measurable instructional objectives to construct qualified test items, determined the objectives of the test before the construction and constructed test drafts with similar mean and SD scores. The SD scores also exhibit as there was no significant difference between the respondents. However, almost the entire test papers exhibit that the test items were not constructed depending on the measurable instructional objectives in the teacher guide of English grade 7. On the other hand, they also signify that English language teachers didn't construct test drafts before the actual tests because the test papers contained many defects that should be avoided.

Item 2 and 4 manifest that English Language teachers almost sometimes categorized instructional objectives in Bloom's taxonomy and prepared ToS to construct tests items with slightly similar mean and SD scores respectively. The SD scores exhibit as there was no significant difference between the respondents.

On the contrary, the test papers which were constructed by the respondents show as they were not constructed depending on the instructional objectives because there were no higher order thinking questions in almost all of the test papers. They also reveal as English language teachers didn't prepare ToS to construct valid tests because a single content is repeatedly asked in different forms. For example, the same item concerning adjectives was asked in different item formats True-false, matching, multiple choice or supply items. Thus, unnecessary repetition of items of the same content confirms as English language teachers didn't prepare ToS. Likewise, Mekonnen et al (2014) reported as ToS are not prepared by all of the school teachers before test construction and teachers simply prepare tests or exams by looking at the major contents of the textbooks ignoring the weight given to each content and such practice of the school teachers signify their test lacks proportionality.

Item 6 designates as teachers usually determined the content area to be tested before constructing the test with 3.41 mean and .78 SD value. The SD score shows as respondents had similar response and becomes low in a homogeneous sample (Dörnyei, 2007). Conversely, the items in the test papers indicate that they were not constructed with adequate sampling because most of them only focus on grammar and vocabulary at large disregarding other language skills. There should be items in the test which directly or indirectly measure students' reading, writing and speaking abilities. The majority of the tests, hence, were not helpful in developing students' language skills in a better way.

Generally, the overall mean score (2.67) shows that teachers almost usually applied the principles of classroom test planning. The overall SD value (1.16) depicts that there was no significant difference between the respondents because the SD values signify as there is no extreme score in the table.

However, the data from the test papers and semi-structured interview indicated that the principles of test planning were not applied as teachers responded in the questionnaire. For instance, the

majority of the test papers contain unnecessary repetition of items of the same content, poor qualities of item formats and inadequate sampling of contents.

One of language department head at CRC School 06 (coded by the researcher) replied that:

“In our department, we don’t bother on the principles of test items construction. Our concern is the score of the students because we are asked why the score becomes low not the instruments which brought the score. So, we use the textbook only. We sometimes use some reference books as well. However, we don’t prepare items based on the table of specifications and lesson objectives. We only take care of whether the content was covered or not.” (15/03/2017 at 1:10 P.M)

Hence, it is possible to deduce that teachers develop test items without adequate planning. Similarly, Siddiek (2010) reported as “It was evident that teachers took very little time in preparing their tests” (p.133).

4.3 Teachers' Response to the Application of Principles of Classroom Test Items Construction

Table 3: Teachers' response to the extent to which they applied the principles of classroom test items construction

№	Items	N	Frequency Scales					M	SD
			4	3	2	1	0		
1	I base each item on an instructional objective.	39	16	18	1	2	2	3.13	1.05
2	I concentrate on the ability to use language for real-life communication.	38	13	17	7	1	0	3.11	.79
3	I determine the test item formats I want to give considering the levels of students.	39	20	10	6	2	1	3.18	1.04
4	I construct a test with different item formats without copying from the textbook.	39	15	11	8	3	2	2.87	1.17
5	I determine the length of a test myself considering the levels such as age and grade of my students.	39	16	15	6	2	0	3.15	.87
6	I determine the difficulty of the items considering the levels such as age and grade of my students.	39	23	7	8	0	1	3.31	.97
7	I avoid items that are easily guessed.	39	7	14	11	3	4	2.44	1.18
8	I use authentic language in my test.	37	14	7	10	2	4	2.68	1.33
9	I avoid constructing tricky items.	39	9	13	13	1	3	2.62	1.11
10	I avoid unintended clues.	39	9	15	11	3	1	2.72	.99
11	I ensure the reliability of my tests constructing test items clearly following the principles.	39	15	14	4	6	0	2.97	1.06
12	I make sure the validity of my test using an adequate sampling of the content taught in class.	39	18	14	5	0	2	3.18	1.02
13	I avoid using contents which reflect bias.	37	21	5	6	3	2	3.08	1.25
14	I write contextualized items.	37	10	11	10	4	2	2.62	1.16
15	I use appropriate wording to the level of education.	39	19	13	4	2	1	3.21	1.00
16	I ensure the readability of my tests by using clear language.	39	25	10	2	1	1	3.46	.91
Grand mean and SD								2.98	1.05

Note: N= Number of respondents who filled the items correctly.

0= Never, **1**= Rarely, **2**= Sometimes, **3**=Usually and **4**= Always

M= Mean **SD**= Standard deviation

Item 1, 3, 12, 13, and 15 in Table 3 present as English language teachers usually based each of the items on instructional objectives of the contents depending on what has been taught in class, determined the test item formats considering the levels of students such as age and grade,

ensured the validity of their test using an adequate sampling of the content taught in class, avoided using contents which reflect bias such as gender, racial, or other grounds, and used appropriate wording to the level of education where the item would be used in tests respectively with almost similar mean and SD scores. The SD scores reveal as the respondents had given similar response.

Nevertheless, the items were not based on the instructional objectives as the data collected from the test papers indicated. The item formats included in the majority of the test papers were also true-false, matching, multiple choice and sometimes supply answer. Even though most item formats were included in the test, the items were very easy; they depended on rote memorization and they didn't encourage students to think critically. The words were also very simple when compared to the words in students' text book. Thus, the tests were below the contents and objectives stated in the students' text book and teachers' guide. The tests were not free from contents which reflect bias such as gender, racial, or other grounds. For example, "_____ is the holy month of Muslims," was a biased item (Appendix J). In addition, the names of people and pronouns in the test papers designate that there was gender and religious bias. On the other hand, the test papers indicated that the tests didn't represent adequate sampling since the items were not constructed using the table of specification.

Item 2, 5, 6, and 16 manifest that English language teachers usually concentrated on the ability to use language for real-life communication when they construct test items, determined the length and difficulty of tests themselves considering the levels such as age and grade of the students and used clear language to construct test respectively with approximate equal mean and SD scores concerning the aforementioned items.

On the contrary, the tests focused on structure and vocabulary. Hence, they didn't concentrate on the ability to use the language for communication even though the students' book exposes students to the use of English language in and out of the classroom environment in a communicative way. The length and difficulty level of the tests were also below the contents and objectives stated in the students' text book and teachers' guide in the majority of the tests. The language used by the teacher was not clear in most test papers as well.

Item 4, 8, 9, 11, and 14 display that English language teachers almost usually constructed tests with different item formats without copying from the textbook, used authentic language, and avoided constructing tricky items. In addition, the items also showed that English language teachers usually ensured the reliability of their test constructing test items clearly following the principle of writing qualified test items formats and constructed contextualized items with slightly similar mean and SD values. The SD scores confirm that there was no significant difference between the respondents.

Conversely, the item formats included in the test papers were true-false, matching and multiple choice and sometimes supply answer. None of the test papers included essay items. Most of the test items were also copied simply changing the names of persons in the textbooks. The language in the test papers, on one hand, was not authentic and the items, on the other hand, were not constructed following the principles of writing qualified test items formats. For instance, the items in matching in most of the test papers were not homogenous; some items were not definite in true-false items; the distractors were not plausible, the alternatives were not listed vertically and not the same in length, the items were decontextualized etc. in multiple choice format.

There was misuse of the contents in reading passage in the textbook as well because they were used to construct test items which were out of the principle of assessment of the target language. For example, “_____ is the largest continent in the world,” in multiple choice items, “The battle of Adwa took place in Adwa on 1st March _____” in the supply items (Appendix J) and “Children miss opportunities in life when they get married early (Appendix N), the government should protect the endemic animals, everyone should discourage early marriage (Appendix P)” and etc in true-false items. The test papers also signify as tricky items were constructed. For example, “Computers do not have brains and we play foot ball in the house” in true-false items (Appendix S) and “Ethiopian coffee is very_____” in supply items (Appendix J) were among tricky items which were constructed by some of English language teachers.

Item 10 designates that teachers nearly usually avoided unintended clues such as grammatical, specific determiners, rote verbal association, length of the correct answer, and length of blank spaces or others to the correct answer with 2.72 mean and .99 SD values. In contrast, the test

papers exhibit that most of the items contained unintended clues, especially, in matching and multiple choice items.

To sum up, the grand mean for the extent to which English language teachers apply the principles of test items construction is 2.98, whereas, the SD score is 1.05. It is possible to deduce that the value of each mean is not too far from the grand mean. Besides, the average SD indicates as there is no significant difference among the respondents when compared with each item's SD. Hence, the data in Table 3 depicts that the principles of test items construction were usually applied among teachers. However, the data from the test papers misrepresent the data gathered from the questionnaires.

Besides, the data gathered from the semi-structured interview of language department heads revealed as the principles of test items construction were not usually applied properly. One of the interviewees in CRC School 04 (coded by the researcher) responded that:

“Some teachers construct test items carelessly. Even they had difficulty of constructing sensible sentences. They also use contents in the reading passage when they construct test items, for example, when students practiced their reading about HIV or Africa, they prepare test items concerning HIV or Africa with the absence of the reading passage which is not acceptable in language testing. I don't believe we are applying the principles all in all. We are told not to include reading passage comprehension items, dialogue items and long items. Generally, we develop 20-30 short items which are printed on a single face of paper due to the shortage of budget for photocopying.” (14/03/2017 at 11:45 A.M)

On the other hand, another interviewee at CRC School 06 replied that:

“We don't prepare essay items because our students' hand writing is not legible. Why do I trouble myself? Even if I develop most students don't attempt to do. So, teachers focus on only objective type of items which easy to score.” (15/03/2017 at 1:10 P.M)

As a result, it is possible to summarize that qualified test items were not prepared following the principles of test construction in language testing.

4.4 Teachers' Response to the Application of the principles of Revising Classroom Tests

Table 4: Teachers' response to the extent to which they revised classroom tests

№	Items	N	Frequency Scales					M	SD
			4	2	2	1	0		
1	I avoid arranging responses in a systematic order or patterns for the sake of scoring items easily.	39	12	14	11	0	2	2.87	1.03
2	I ensure that each item is independent of all other items.	39	17	11	9	1	1	3.08	1.01
3	I ensure that there is a single clearly correct (or best) answer for each item.	39	22	9	6	1	1	3.28	.99
4	I proofread my test to avoid errors such as grammatical, spelling and typing before reproducing it.	39	24	8	6	0	1	3.38	.93
5	I arrange test items from easy to difficult.	39	18	13	6	1	1	3.18	.97
6	I arrange test item formats in appropriate order.	39	11	13	9	4	2	2.69	1.15
7	I give my test to other colleagues to help me examine and make comments.	38	8	11	13	4	2	2.50	1.10
8	I consider the comments and criticism offered by other colleagues in order to improve the quality of my test.	39	10	13	12	3	1	2.72	1.02
Grand mean and SD								2.96	1.02

Note: N= Number of respondents who filled the items correctly.

0= Never, **1**= Rarely, **2**= Sometimes, **3**=Usually and **4**= Always

M= Mean **SD**= Standard deviation

Item 1, 6, 7 and 8 depict that teachers almost usually avoided arranging responses in a systematic order or patterns for the sake of scoring items easily, arranged test item formats in appropriate order, gave their test to other colleagues to help them examine and make comments and considered the comments and criticism offered by other colleagues in order to improve the quality of their tests respectively with approximately equivalent mean and SD scores. The SD scores exhibit that there was alike response between the sample teachers.

Likewise, the data gathered from the test papers reveal as English language teachers avoided arranging responses in a systematic manner for the sake of marking items easily. Item formats were written in appropriate order as well.

In contrast, the data gathered from the test papers and semi-structured interviews disclose as English language teachers didn't give their test to other colleagues to help them examine and make comments. Even if the teachers were given comments, they didn't consider the comments and criticism offered by other colleagues in order to improve the quality of their tests. Consequently, there is no culture of commenting tests among English language teachers because all of the test papers contained so many shortcomings.

Item 2 shows that teachers usually ensured that each item is independent of all other items with 3.08 mean and 1.01 SD scores. Conversely, the items in most test papers demonstrate as there were dependent items in a test paper. For example, a single item about adjective in true-false item was repeatedly asked in multiple-choice or in supply item in the different form.

Item 3, 4 and 5 signal that English language teachers usually ensured that there is a single clearly correct (or best) answer for each item, proofread their tests to avoid errors such as grammatical, spelling and typing before reproducing them and arranged test items from easy to difficult respectively with more or less equivalent mean and scores. The SD scores exhibit that there was no significant disparity the response of the sample teachers.

The test papers, however, signify as some English language teachers didn't check that there were a single correct answer and ambiguous items. The test papers were not also proofread properly before reproducing them because all of the test papers contained grammatical, spelling and typing errors. This, in turn, caused test administration problems. The test items were not arranged from easy to difficult as well.

In general, the grand mean for the application of principles of revising classroom tests is 2.96, whereas the SD score is 1.02. It is possible to infer that the value of each mean is not too far from the grand mean. Besides, the average SD indicates as there is no significant variation among the respondents when compared with each item's SD in table 4. Hence, the data portray that the principles of revising test items were usually applied among teachers.

However, the data gathered from the test papers and interviews falsify the data gathered from the questionnaires. The test papers demonstrated that the test items were not revised because the test papers contained incomplete items, wrong items, grammatical and spelling errors. Pertaining to

this, most of the interviewees reported that they evaluated the constructed test items at their department but they complained that the aforementioned problems occurred when the items were typed by the computer. They claimed that test papers were reproduced without being edited by responsible teachers. However, most department heads responded that there was no trend of giving tests to colleagues to be commented. Thus, the data portrayed that there were problems in revising the test items once they had been constructed.

4.5 Triangulation of Data in Planning, Constructing and Revising Classroom Tests

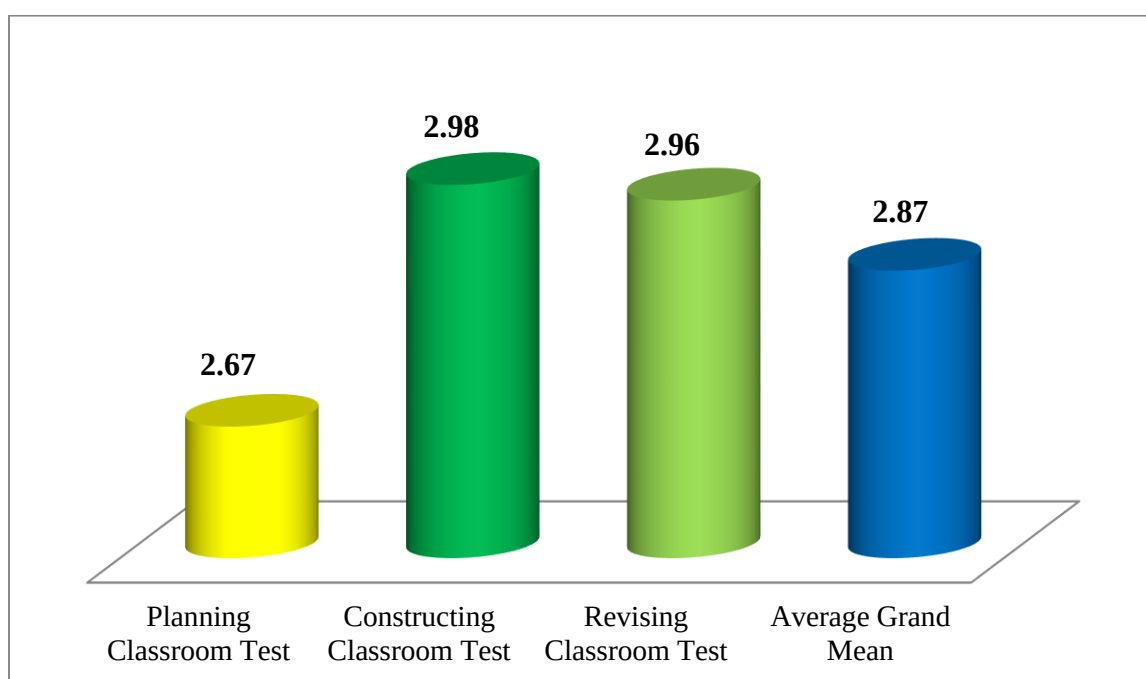


Figure 1: Overall mean scores in English language test items construction

As Figure 1 indicates, the data gathered from the questionnaire show that English language teachers applied the principles of planning, constructing, and revising classroom tests with grand mean 2.67, 2.98, and 2.96 respectively. The overall average grand mean score is 2.87. This signifies teachers usually applied the principles. However, the data gathered from semi-structured interview and document analysis proved as the principles were not properly applied. Hence, the overall cross-checked result shows that English language teachers prepared test items without adequate planning and considerations of the principles of test items construction and revising.

4.6 Teachers' and Students' Response to the Application of Pre-test Administration Principles

Table 5: Teachers' response to the extent to which they applied principles of pre-test administration

№	Items	N	Frequency					M	SD
			4	3	2	1	0		
1	I prepare my test long before its administration so that it is not done in a hurry.	39	8	14	13	3	1	2.64	.98
2	I announce a test at least a week before.	39	12	11	12	2	2	2.74	1.11
3	I inform my students about the time of test administration.	39	14	15	10	0	0	3.10	.78
4	I check for the readiness of all materials for test administration.	39	20	13	4	2	0	3.31	.86
5	I remind students to get to the classroom early.	38	22	8	6	1	1	3.29	1.01
6	I give my students practice tests.	39	20	10	4	1	4	3.05	1.29
7	I identify the negative markers: underline, type them in capital letters or put them in boldfaced type.	39	8	16	14	0	1	2.77	.87
8	I inform the students what types of questions will be on the test (essay, multiple-choice, etc.) priorly.	39	15	12	6	5	1	2.90	1.14
9	I indicate how much the examination will count toward determining the final grade.	39	13	17	8	1	0	3.08	.80
10	I avoid constructing time-consuming tests.	37	10	20	3	4	0	2.97	.89
11	I write clear instructions.	39	24	9	5	1	0	3.44	.82
12	I keep test questions securely.	38	23	10	4	0	1	3.42	.88
13	I prepare enough copies of the test papers.	39	23	13	2	1	0	3.49	.72
14	I reproduce aesthetically pleasing copies.	39	14	14	9	1	1	3.00	.97
15	I avoid administering too long tests that bring administration difficulties like test anxiety.	39	17	14	7	1	0	3.21	.83
Grand mean and SD								3.09	.93

Note: N= Number of respondents who filled the items correctly.

0= Never, **1**= Rarely, **2**= Sometimes, **3**=Usually and **4**= Always

M= Mean **SD**= Standard deviation

Table 6: Students' response to the extent to which their English teachers applied principles of pre-test administration

№	Items	N	Frequency					M	SD
	Our English teacher:		4	3	2	1	0		
1	announces a test at least a week before.	88	28	25	18	14	3	2.69	1.17
2	informs us about the time of test administration.	88	43	15	13	9	8	2.86	1.36
3	checks for the readiness of all materials for test administration.	87	44	21	14	4	4	3.11	1.12
4	reminds us to get to the classroom early.	86	54	16	5	3	8	3.22	1.27
5	gives us practice tests before the actual tests.	86	38	14	9	10	15	2.58	1.56
6	identifies the negative markers: underline, type them in capital letters or put them in boldfaced type.	87	38	19	13	9	8	2.80	1.34
7	informs us what types of questions will be on the test prior to the examination.	88	21	13	15	19	20	1.95	1.50
8	indicates to us how much the examination will count.	87	49	8	13	11	6	2.95	1.36
9	arranges proper sitting arrangement.	87	39	21	9	9	9	2.83	1.37
10	administers a test at the proper time of the day.	88	14	16	17	11	30	1.69	1.49
11	writes concise instructions.	88	55	21	5	3	4	3.36	1.05
12	keeps test questions securely.	86	62	11	4	4	5	3.41	1.15
13	prepares enough copies of the test papers.	85	55	18	5	4	3	3.39	1.03
14	ensures that there is enough light in the classroom.	87	28	20	14	12	13	2.44	1.44
15	avoids administering too long tests.	88	27	14	23	11	13	2.35	1.41
16	reproduces aesthetically pleasing copies.	87	46	19	9	6	7	3.05	1.28
17	uses different types of test item formats.	87	55	14	11	6	1	3.33	1.01
18	constructs tests free from grammatical errors.	88	22	20	16	25	5	2.33	1.28
19	constructs tests free from spelling errors.	88	27	23	17	14	6	2.59	1.27
20	avoids constructing tricky items.	88	16	13	20	15	24	1.80	1.45
Grand mean and SD								2.73	1.29

Note: N= Number of respondents who filled the items correctly.

0= Never, **1**= Rarely, **2**= Sometimes, **3**=Usually and **4**= Always

M= Mean **SD**= Standard deviation

Item 1, 7 and 10 in Table 5 demonstrate as English language teachers nearly usually prepared their test long before its administration, identified the negative markers, type them in capital letters or put them in boldfaced type and avoided constructing time-consuming tests respectively with related mean and SD scores. The SD scores illustrate the respondents were homogenous in their response. Conversely, the data collected from the test papers corroborate that teachers

didn't take enough time to construct tests because almost all of the papers disclose as they were constructed haphazardly. The negative markers were not also written in neither capital nor boldfaced type. On the other hand, the students' response in table 6 designates as their English language teachers sometimes avoided time consuming tests with 2.35 mean score and 1.41 SD value. The mean scores substantiate that there is a minor disparity between teachers' and students' response but the SD scores affirm that there is considerable disparity among students more than teachers.

Item 2 and 8 in Table 5 disclose that English language teachers usually announced a test at least a week before and informed the students what types of questions will be on the test (essay, multiple-choice, etc.) prior to the examination correspondingly with almost similar mean and SD values. The SD scores point up that the respondents were identical in their response. The students' response on Table 6 also shows as their teachers usually announced the test a week before with 2.69 mean score and 1.17 SD value. These data confirm that there is no significant difference between teachers' and students' response. In addition, the data from the observation corroborates as students were informed before the test from the notice boards in the sample schools. On the contrary, the students' response signifies as they were sometimes informed about test item formats with 1.95 mean score and 1.5 SD value. The means and SD depict that there is a considerable difference between the respondents.

Item 5 and 6 in Table 5 portray that English language teachers usually reminded students to get to the classroom early and gave their students practice tests with almost parallel mean and SD scores. The SD scores indicate that there was no significant disparity between the respondents concerning the two items. Similarly, the students' response shows as teachers usually reminded them about the time of the test with 3.29 mean score and 1.27 SD value. These data confirm that there is no significant difference between teachers' and students' response. Conversely, the data from the observation disclose that some of the students were late for the exam and even very few students arrived when students had finished doing the exam. The students' response indicates as they were usually given practice tests with 3.58 mean score and 1.56 SD value. The means and SD prove that there is a slight difference between teachers' and students' response.

Item 3, 4, 9, 11, 12, 13, 14 and 15 in Table 5 present that English language teachers usually carried out the following activities: informing students about the time of test administration, checking for the readiness of all materials for test administration, indicating the weight of the test, writing clear instructions, keeping test questions securely, preparing enough and aesthetically pleasing copies of the test papers, and avoiding administration of too long tests that bring administration difficulties like test anxiety respectively nearly with the same mean and SD scores. The SD scores disclose that the respondents were more homogenous in their response than the other items in the table.

Likewise, the students' response in Table 6 shows as they were usually aware of the time of testing and indicated the weight of the test with similar mean and SD values. The means reveal that there is similarity of response but the SD values exhibit that there is a slight difference between teachers' and students' response. The data from the observation also affirm as students were informed the time of the tests from the notice boards in the sample schools but the weight of the test was neither told nor written on the test papers.

Correspondingly, the students' response manifests as their English language teachers usually checked the readiness of materials before the administration of tests and wrote clear instructions for the items with almost parallel mean and SD values. The means connote that there is a resemblance of responses but the SD values display that there is a slight difference between teachers' and students' response. On the contrary, unclear instructions were seen in the test papers. For example, say true or false, match A with B, etc. There were also some problems such as shortage of students' seat and classroom, scarcity of test papers and crowded and dirty classrooms in the readiness of test administration in the sample schools

The students' response in Table 6 also conveys as enough copies were usually prepared and the copies of test papers were aesthetically pleasing with relative mean and SD scores. These data substantiate that there was no significant difference between teachers' and students' response but the SD scores verify that there is considerable disparity among students more than teachers. Nevertheless, the data collected from the observations and interviews corroborate that there was a shortage of copies. Teachers were busy in collecting excess copies from other classes. Therefore the distribution of the test becomes unfair. The copies of the majority of test papers, on

the other hand, were aesthetically pleasing except that of Appendix R in which the items were not clearly visible.

To sum up, the grand mean for the application of the principles of pre-test administration is 4.09, whereas the SD score is .93 for teachers and 2.73 mean and 1.29 SD score for students. It is possible to decide that the value of each mean is not too far from the grand mean. Besides, the average SD indicates as there is a slight difference among the respondents when compared with each item's SD. Thus, the average scores of the data show that English language teachers almost usually applied the principles of pre-test administration. However, the data gathered from the semi-structured interviews and observations to some extent misrepresent the data gathered from the questionnaire.

4.7 Teachers' and Students' Response to the Application of While-test Administration Principles

Table 7: Teachers' response to the extent to which they applied principles of while test administration

No	Items	N	Frequency					Mean	SD
			0	1	2	3	4		
1	I prefer to administer my test myself.	39	13	18	7	0	1	3.08	.87
2	I distribute test papers properly.	39	24	7	6	1	1	3.33	1.01
3	I remind my students to mentally figure out how much time they will need for each part.	39	13	14	10	2	0	2.97	.90
4	I inform students to look over the whole test quickly in order to get a good grasp of its different parts after the test papers are distributed.	39	10	12	11	5	1	2.64	1.08
5	I administer my test considering conducive environment such as light, sitting arrangement, weather condition, etc.	39	16	17	4	2	0	3.21	.83
6	I consider my students' psychological conditions.	39	14	13	9	3	0	2.97	.95
7	I ensure proper supervision while administering my test	38	17	12	9	0	0	3.21	.81
8	I handle any case of exam malpractices (e.g. cheating) in a regulated way.	39	16	16	6	1	0	3.21	.80
9	I advise my students to concentrate on the test as carefully as possible.	39	23	11	3	1	1	3.38	.93
10	I allow sufficient time for test completion.	38	18	9	8	3	0	3.11	1.00
11	I inform students a few minutes before the end of test time so that they can finish on time, proofread their answers, and catch errors.	39	23	9	5	2	0	3.36	.90
Grand mean and SD								3.13	.91

Note: N= Number of respondents who filled the items correctly.

0= Never, **1**= Rarely, **2**= Sometimes, **3**=Usually and **4**= Always

M= Mean **SD**= Standard deviation

Table 8: Students' response to the extent to which their English teachers applied principles of while test administration

No	Items	N	Frequency					M	SD
	Our English teacher:		4	3	2	1	0		
1	distributes test papers properly.	87	63	13	7	2	2	3.53	.91
2	reminds us to mentally figure out how much time we will need for each part of the test paper.	88	32	21	9	10	16	2.49	1.52
3	advises us to concentrate on the test.	87	22	25	26	11	3	2.60	1.10
4	informs us a few minutes before the end of the test time.	88	36	26	11	10	5	2.89	1.22
5	prefers to administer his/her test him/herself.	86	32	18	9	7	20	2.41	1.60
6	informs us to look over the whole test quickly.	88	29	20	20	8	11	2.55	1.36
7	makes proper supervision.	88	42	23	10	10	3	2.03	1.16
8	handles any case of exam malpractices (e.g. cheating) in the regulated way.	88	36	21	9	14	7	2.75	1.35
9	allows us sufficient time for test completion.	88	55	12	12	5	4	3.24	1.16
10	considers our psychological conditions.	85	31	25	15	7	7	2.78	1.25
11	administers tests teaching us the contents.	86	49	13	8	4	12	2.97	1.45
Grand mean and SD								2.84	1.28

Note: N= Number of respondents who filled the items correctly.

0= Never, **1**= Rarely, **2**= Sometimes, **3**=Usually and **4**= Always

M= Mean **SD**= Standard deviation

Item 1, 5, 7, 8, 9, and 11 in Table 7 connote as English language teachers usually performed the following while administering the tests they had constructed: preferring to administer their test by their own, administering their tests considering conducive environment such as light, sitting arrangement, weather condition, etc., making proper supervision, handling any case of exam malpractice (e.g. cheating) in the regulated way, advising their students to concentrate on the test as carefully as possible and informing students a few minutes before the end of test time so that they can finish on time, proofread their answers, and catch errors in respecting items with relative mean and SD scores. The SD scores in the listed items are lower than the other scores so that the respondents are homogeneous in their response.

Similarly, the data collected from the students indicates that English language teachers usually made proper supervision, handled any case of exam malpractices (e.g. cheating) in the regulated

way, advised their students to concentrate on the test as carefully as possible and informed students a few minutes before the end of test time so that students can finish on time, proofread their answers, and catch errors with almost similar mean and SD scores. The data collected from the interviews, observations, and open ended questions in students' questionnaire, on the contrary, corroborates as there was no proper supervision, exam malpractices were not properly handled, students were not advised to concentrate on the test as carefully as possible and students were not informed about their time periodically.

The students' response in Table 8, on the other hand, points out as their English language teachers almost sometimes administered their own tests considering conducive environment such as light, sitting arrangement, weather condition, etc. In contrast, the data collected from the observation confirm as there were environmental problems such as shortage of light in some classes, poor sitting arrangement, shortage of seats, noise around classrooms etc.

Item 2 and 10 in Table 7 disclose as English language teachers usually distributed test papers properly and allowed sufficient time for test completion respectively with almost similar mean and SD values. The SD values indicate as there is no considerable disparity between the respondents. Likewise, the data gathered from the students' response illustrates as English language teachers usually performed the aforementioned activities with similar mean and SD scores. However, the data collected from the observations denote as there was no fair distribution of tests because students in some classes got the test papers and started working on the test items earlier than some other classes. Even though the time was not mentioned in the test papers and students were not informed, it was observed that sufficient time was given to the students.

Similarly, one of the interviewees at CRC School 07 (coded by the researcher) replied that:

"We don't let our students do tests within the given time. Tests are not administered within similar time at schools of our CRC as well. Even time is not limited. Students return their test papers when they finish working on the test by themselves. There is no trend of informing time in an interval." (15/03/2017 at 2:45 P.M)

Item 3 and 6 in Table 7 signal as English language teachers nearly usually reminded their students to mentally figure out how much time they would need for each part and considered

their students' psychological conditions correspondingly with almost similar mean and SD values. The SD values designate that the respondents are similar in their response. The students' response in Table 8 portrays as their English language teachers usually reminded them to mentally figure out how much time was needed to work on each part of the item formats and considered the students' psychological conditions. Nevertheless, the data collected from the observations and interviews affirm as nothing was told regarding the time and the students' psychological conditions were undermined in some classes because some teachers were aggressively administering the test.

Item 4 conveys as English language teachers almost usually informed students to look over the whole test quickly in order to get a good grasp of its different parts after the test papers are distributed with 2.64 mean and 1.08 SD values. The students response, whereas, shows that they were sometimes informed to look over the whole test quickly. In contrast, the data gathered from the observations and interviews manifest that students were not informed to look over the whole test.

By and large, the grand mean for the application of principles of while test administration is 3.13, whereas the SD score is .91 for teachers and 2.84 mean and 1.28 SD value. It is possible to decide that the value of each mean in Table 7 and 8 is not too far from the grand means. Furthermore, the average SD indicates as there was no considerable difference among the respondents when compared with each item's SD value in both tables. Hence, the average scores of the data show that English language teachers usually applied the principles of while test administration. However, the data from the class observations and interviews misrepresent the data gathered from the questionnaires. So, some English language teachers had poor experiences of administering tests because of the problems observed during the administration of the test that harm the reliability of the test.

4.8 Teachers' and Students' Response to the Application of Post-test Administration Principles

Table 9: Teachers' response to the extent to which they applied principles of post-test administration

№	Items	N	Frequency					M	SD
			4	3	2	1	0		
1	I collect answer sheets properly.	39	28	6	3	0	2	3.49	1.02
2	I encourage my students to discuss to evaluate the quality of the test items immediately.	39	16	11	7	3	2	2.92	1.17
3	I give immediate feedback to my students.	39	17	10	10	0	2	3.03	1.08
4	I encourage my students to comment on the general test administration procedures.	39	15	7	11	2	4	2.69	1.32
Grand mean and SD								3.03	1.14

Note: N= Number of respondents who filled the items correctly.

0= Never, 1= Rarely, 2= Sometimes, 3=Usually and 4= Always

M= Mean SD= Standard deviation

Table 10: Students' response to the extent to which their English teachers applied principles of post-test administration

№	Items	N	Frequency					M	SD
			4	3	2	1	0		
1	Collects our answer sheets fairly.	87	67	11	5	4	0	3.62	.79
2	Encourages us to discuss to evaluate the quality of the test items immediately.	85	20	24	11	15	15	2.22	1.44
3	Gives us immediate feedback.	86	30	16	12	15	13	2.41	1.49
4	Encourages commenting on the general test administration procedures.	88	17	16	14	13	28	1.78	1.53
Grand mean and SD								2.50	1.31

Note: N= Number of respondents who filled the items correctly.

0= Never, 1= Rarely, 2= Sometimes, 3=Usually and 4= Always

M= Mean SD= Standard deviation

Item 1 and 3 in Table 9 signify as English language teachers usually collected answer sheets properly from students and gave immediate feedback to their students with almost similar mean and SD values. The SD values indicate as the respondents are similar in their response. The students' response in Table 10 shows that their English language teachers usually collected answer sheets properly from students parallel to teachers with almost alike mean values but the SD values exhibit students are more homogeneous than teachers in responding to the item. However, students reported that they sometimes got immediate feedback from teachers in Table 10 and open-ended questions.

Item 2 and 4 in Table 9 demonstrate as English language teachers about usually encouraged their students to evaluate the quality of the test items immediately and to comment on the general test administration procedures with almost similar mean and SD values. The SD values indicate as the respondents are similar in their response. Conversely, the students' response in Table 10 reveal as they were sometimes encouraged to evaluate the quality of the test items immediately and to comment on the general test administration procedures with different mean and almost similar SD values.

On the whole, the average mean score for the application of the principles of test administration is 3.03, whereas the SD score is 1.14 for English language teachers. The value of each mean is not far from the grand mean. Hence, teachers usually apply the principles of post-test administration. Furthermore, the average SD indicates as there was no significant difference among the respondents when compared with each item's SD. On the contrary, the average mean score for the application of the principles of test administration is 2.5, whereas the SD score is 1.31 for students. The average SD indicates as there was a slight difference among the respondents. The students were more homogeneous in responding to item 1 in Table 10 than the other items. Nevertheless, the data gathered from the interview falsify the data gathered from the questionnaires. One of the interviewees at CRC School 05 (coded by the researcher) replied that:

“Actually, we don't let students comment on the test items because they can't. They don't ask us even if we give them wrong items. We don't also evaluate the administration process with teachers in our department. We bother on reporting students' scores to the school directors.”
(15/03/2017 at 9:30 A.M)

It is possible to infer that there is no trend of evaluating test items and the administration process within the department teachers and students in classes.

4.9 Triangulation of Data in Pre-test, While-test and Post-test Administration

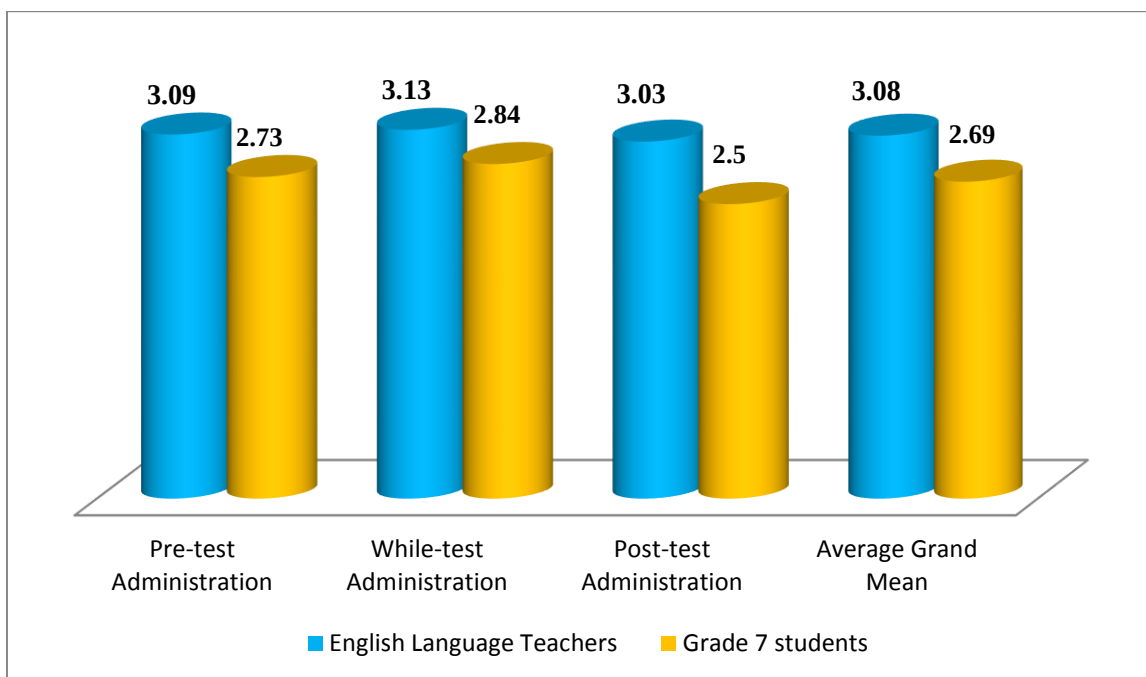


Figure 2: Overall mean scores in English language test items Administration

As shown in Figure 2, the data gathered from English language teachers' questionnaire show that they applied the principles of pre-test, while-test and post-test administration with mean 3.09, 3.13, and 3.03 respectively, whereas, the data from the students' questionnaire indicate English language teachers applied the principles with mean 2.73, 2.84 and 2.5 respectively. The overall average grand mean score for teachers and students is 3.08 and 2.69 respectively. The scores point out as English language teachers usually applied the principles of test administration. Conversely, the data gathered from semi-structured interview and classroom observations proved as the constructed tests were not administered according to principles although English language teachers seemed to know them. Therefore, the overall crosschecked result shows that English language teachers didn't administer the test items they prepared considering the principles of test administration.

Chapter Five: Summary of the Findings, Conclusions and Recommendations

This chapter consists of the summary of the findings, general conclusions and recommendations which are assumed to improve the practical applications of English language test items construction and administration in the second cycle of primary schools of Jeju Woreda.

5.1 Summary

Testing and teaching are interconnected phenomenon (Desheng & Varghese, 2013). Consequently, the Ethiopian education policy favors the importance of testing as one the techniques of assessment in the education system (MoE, 2002). Language testing differs from testing in other fields. This means it needs careful preparation.

Similarly, the purpose of this study was to analyze the practical applications of the principles of English language test items construction and administration in the second cycle of primary schools of Jeju Woreda. To meet this purpose, the following basic research questions were developed.

1. To what extent do teachers apply the principles of test items construction?
2. What are the shortcomings of teacher-made English language test items?
3. How are the constructed tests administered in the study area?

To answer these questions, the researcher employed a descriptive survey study. English language teachers, teacher-made English language test papers, language department heads, grade 7 students and English classes were the target population of the study. Thus, 45 English language teachers who were teaching English in the SCPSs of Jeju Woreda were included in the study using census sampling. Whereas, 13 teacher-made English language test papers, 8 language department heads, 88 grade 7 students and 7 English classes were chosen using simple random sampling technique.

The data were obtained through questionnaires, semi-structured interview, semi-structured observation and document analysis. Pertaining to these, questionnaires were distributed to English language teachers and grade 7 students. As a result, 4 questionnaires were not returned

and 2 of them were disqualified because they were carelessly filled. A semi-structured interview was also conducted with 8 department heads in CRC represented schools. 7 English classroom observations were carried out several times to collect data on how teachers administer tests in the sample schools as well. Moreover, 13 English language test papers from the 13 CRC schools were collected depending on test evaluation techniques.

The data gathered through closed-ended questionnaire were coded, tabulated and analyzed using descriptive statistics such as frequency, mean and standard deviation by using SPSS version 20 software. The additional data, on the other hand, gained from semi-structured interview, document analysis and English class observations were also coded, categorized and identified thematically in order to approve or disapprove the quantitative data. The following major findings were discovered from the result of data analysis.

i. Basic Research Question 1: To what extent do teachers apply the principles of test items construction?

The average grand means in Table 2, 3 and 4 concerning the applications of the principles of test items construction was 2.84. This indicated as English language teachers applied the principles almost usually. They also recognized what happens when test items are not constructed following their guiding principles as the data collected from the open-ended questions revealed. However, the data collected from the interview and document analysis disproved the data collected from the teachers' questionnaire.

Similar to Mekonnen et al (2014) and Teshome's (2015) finding, English language teachers didn't prepare ToS to construct valid tests. They also didn't take enough time to construct tests because almost all of the papers disclosed as they were constructed haphazardly without taking into account the principles of English language test items construction. Besides, teachers didn't revise and comment each other on test items because the test papers contained defects such as ambiguous, incomplete, wrong and biased items, grammatical and spelling errors. Look at Appendix H-T.

Generally, it is possible to deduce that English language teachers developed test items without adequate planning and considering the principles of test items construction.

ii. Basic Research Question 2: What are the shortcomings of teacher-made English language test items?

Teacher-made achievement tests have poor quality (Mehrens & Lehmann, 1991). English language teachers developed test items without adequate planning and considering the principles of test items construction. Test items developed in such manner are not free from defects. Likewise, test items constructed by English language teachers contained the following serious shortcomings.

- The tests were not valid because they contained unnecessary repetition of items of the same content.
- The tests depended on rote memorization and they didn't encourage students to think critically.
- The length and difficulty of the tests were below the level of contents and objectives stated in the students' text book and teachers' guide.
- The tests focused on structure and vocabulary only. Other language skills were disregarded.
- The language in the test papers was not authentic and didn't concentrate on the ability to use the language for communication.
- There was a misuse of the contents in the reading passage in the textbook because they were used to construct test items which were out of the principle of assessment of the target language.
- The test papers contained decontextualized and dependent items.
- The test papers contained ambiguous, incomplete, wrong and biased items, grammatical, spelling and typing errors because they were not proofread properly before they were reproduced.

iii. Basic Research Question 3: How are the constructed tests administered?

The average grand means in Table 5, 7 and 9 about the applications of the principles of test administration for teachers' questionnaire was 3.08. This exhibited that English language teachers applied the principles almost usually. Similarly, the data obtained from the students' questionnaire was in favor of the result with average means 2.69 in Table 6, 8 and 10.

On the contrary, the data gathered from the interview and classroom observation showed that the constructed tests were not administered according to the principles of test administration. The interviewees replied that teachers in their department were aware of the principles but they didn't usually apply them well. In addition, the data gathered from the class observation indicated that there was no proper supervision and good time management; exam malpractices were not properly handled; students were not advised to concentrate on the test as carefully as possible and students were not informed about their time periodically. Besides, there were environmental problems such as shortage of light in some classes, poor sitting arrangement, shortage of seats, noise around the classrooms, etc.

5.2 Conclusions

Based on the findings of the study, the following conclusions were drawn:

English language teachers were aware of the principles of test items construction; however, they developed test items without adequate planning taking no notice of the principles of test items construction.

Furthermore, English language test items contained serious defects because they were prepared without adequate planning and the considerations of the principles of test items construction. For instance, there were ambiguous, incomplete, wrong and biased items, and grammatical, spelling and typing errors because they were not proofread properly before they were reproduced. The tests items also focused on structure and vocabulary only taking no notice of other language skills.

Lastly, the constructed tests were not administered according to the principles of test administration. Even though English language teachers knew the principles, they didn't usually apply them. Consequently, there was no proper supervision and good time management; exam malpractices were not properly handled; students were not advised to concentrate on the test as carefully as possible and students were not informed about their time periodically. There were also environmental problems such as shortage of light in some classes, poor sitting arrangement, shortage of seats, noise around the classrooms, etc.

5.3 Recommendations

Based on the findings of the study, the following recommendations were forwarded:

1. English language teachers developed test items without adequate planning although they were aware of the principles of test items construction. Therefore, English language teachers should take time to prepare test items considering the principles of test items construction. On the other hand, the constructed tests were not administered properly. Hence, English language teachers should administer their tests according to the principles of test administration. Experience sharing program should also be held in language departments of the schools on the basis of the principles of test items construction and administration.
2. The study found out that English language test items which were prepared by English language teachers contained serious defects. Thus, contextualized training should be given to English language teachers on how to construct a qualified English language tests based on the defects they frequently make. Moreover, colleges of teacher training should train teachers on how to construct and administer tests properly.
3. Language department heads and English language teachers in the open-ended items explained that they were enforced to prepare very short test items with inadequate content sampling; ignoring reading, speaking and writing. They were also ordered that the items should be completed on a single face of paper because of the shortage of budget to reproduce tests. This shows that the importance of testing is not equally considered as teaching. Hence, English language teachers should not be limited to prepare qualified tests due to budget restriction. The schools should spend enough budgets as much as teaching.

Recommendations for Further Study

This study tried to analyze the current practice of English language test items construction and administration in SCPSs of Jeju Woreda. There are still some areas untouched in the topic. Thus, the following are suggestions for further studies:

- 🍏 The study focused on all SCPSs in the Woreda. Therefore, it is possible to conduct a comparative study on the practice of English language test items construction and administration between CRC schools.
- 🍏 The study was limited to government SCPSs in a single Woreda. Thus, it would have been more comprehensive and general if it had been conducted at Zone level.
- 🍏 The study was limited to government SCPSs. Hence, it is possible to conduct a comparative study on the practice of English language test items construction and administration in private and government schools.
- 🍏 The study was limited to English language. So, it is possible to conduct similar study in other subjects.

References

- Abatihun Alehegn. (2013). Instructors' Classroom Assessment Practices as a Function of Training Background and Teaching Experience with Particular Reference to Final Examinations. *American Journal of Educational Research* 1(8). P.300-306. Retrieved on September 15, 2013 from <http://www.pubs.sciepub.com>.
- Alade, O.M. & Omoruyi, I. V. (2014). Table of Specification and Its Relevance in Educational Development Assessment. *European Journal of Educational and Development Psychology*. 2 (1) pp.1-17. Retrieved March from <http://www.ea-journals.org>.
- Alderson, J. C, Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: CUP.
- Ashenafi Tilahun. (2014). The Practices and Challenges of Test Construction and Administration in Selected Secondary Schools of Kolfe Keranio Sub-City in Addis Ababa. Unpublished thesis. Retrieved on 25 August from <http://www.AAU.et>
- Atkins, J., Hailom, B. & Nuru, M. (1996). *Skills Development Methodology*. Addis Ababa: Addis Ababa University Printing Press.
- Bachman, (1990). *Fundamental Considerations in Language Testing*. Oxford: OUP.
- Bachman, L. F. & Palmer A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Tests Language Tests*. Oxford: Oxford University Press.
- Bachore, M. M. (2015). The Status, Roles and Challenge of Teaching English Language in Ethiopian context: The case of selected primary and secondary schools in Hawassa University Technology Village Area. *International Journal of Sociology Education* , 4(2) P. 182-196. Retrieved on November 10, 2015 from <http://rise.hipatiapress.com>
- Borich, G. & Kubiszyn, T. (2013) *Educational Testing and Measurement: Classroom Application and Practice*. 10th ed. Hoboken: John Wiley & Sons, Inc.
- Broughton, G., Brumfit, C., Flavell, R., Hill, P. & Pincas, A. (1980). *Teaching English as a Foreign Language*. 2nd ed. London: Routledge.

- Brown, H.D. (2004). *Language Assessment: Principles and Classroom Practices*. White Plains, NY: Pearson.
- Brown, J. D. & Hudson, T. (1998). The Alternatives in Language Assessment. TESOL QUARTERLY 32(4) p.653-675. Retrieved in Winter 1998 from <http://www.personal.psu.edu>
- Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall.
- CELA. (2016). Principles of Good Practice: Research and innovation in language learning and assessment. Retrieved on 22/10/2016 from <http://www.cambridgeenglish.org>
- Ch'ng, L. & Rethinasamy, S. (2013). English language assessment in Malaysia: Teachers' practices in test preparation. Issues in Language Studies 2(2) p. 24-39. Retrieved on 22/10/2016 from <http://ir.unimas.my>
- Clay, B. (2001). *Is This a Trick Question? A Short Guide to Writing Effective Test Questions*. Kansas: Kansas Curriculum Center. Retrieved in October 2001 from <http://www.k-state.edu>.
- Desheng, C. & Varghese, A. (2013). Testing and Evaluation of Language Skills. IOSR Journal of Research & Method in Education. 1(2). PP. 31-33. Retrieved in March to April, 2013 from <http://www.iosrjournals.org>
- Ebel, R. L. & Fresbie, D. A. (1991). *Essentials of Educational Measurement*. 5th ed. Englewood Cliffs: Prentice Hall.
- Fives, H. & DiDonato-Barnes, N. (2013). Classroom Test Construction: The Power of a Table of Specifications. Practical Assessment, Research & Evaluation, 18(3). Available online: <http://pareonline.net/getvn.asp?v=18&n=3>
- Fulcher, G. & Davidson, F. (2007). *Language Testing and Assessment: An Advanced Resource Book*. Milton Park: Routledge.
- Fulcher, G. (2010). *Practical Language Testing*. London: Hodder Education.

- Gray, D. E. (2004). *Doing Research in the Real World*. London: SAGE Publications Ltd.
- Harmer, J. (2007). *how to teach English*. Edinburgh Gate: Pearson Education Limited.
- Heaton J. B. (1990). *Writing English Language Tests*. New York: Longman inc.
- Hughes, A. (2003). *Testing for Language Teachers 2nd ed*. Cambridge: CUP.
- ICDR. (2002). *Teacher Education Handbook*. 2nd ed. Addis Ababa: FinFinne Printing and Publishing S.C.
- Jeju Woreda Communication office. (2016/17). Baga Nagaan Dhuftan! Aanaa Jajuu Beektuu? [Welcome! Do You Know Jeju Woreda?] A monthly brochure.
- Kolawole, C.O.O. (1998). Problems of Tests Constructed by Language Teachers in Secondary schools in Ondo State. *Journal of Educational Research and Evaluation* (Vol. 2) pp.22-25.
Retrieved on 15/11/2016 from <http://www.iiste.org>
- Kothari, C. R. (2004). *Research methodology: Methods and Techniques*. New Delhi: New Age International (P) Ltd.
- Lewin, C. (2005). Elementary Quantitative Methods. In Somekh, B. & Lewin, C. (Eds.), *Research methods in the social science* (p215-225). London: SAGE Publications Ltd.
- Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2006) *Methods in Educational Research: From Theory to Practice*. San Francisco: Jossey-Bass.
- Lodico, M. G., Spaulding, D. T., & Voegtle, K. H. (2010) *Methods in Educational Research: From Theory to Practice. 2nd ed*. San Francisco: Jossey-Bass.
- Madsen, H. S (1983) *Techniques in Testing*. Oxford: OUP.
- Mehrens, W. A. & Lehmann, I. J. (1991). *Measurement and Evaluation in education and psychology*. USA: Wadsworth.

- Mekonnen Yibrah. Dr. Julia D. & Dr. Deepika N. (2014). Assessment of Content Validity in Ethiopian General Secondary Education Certificate of English Examinations (2009-2012). *Global Scholars Journal of Linguistics and Communication Studies* 1 (1), P 1-13. Retrieved May, 2014 from <http://www.globalscholarsjournals.org>
- Meseret Kuma. (2011). Test Construction Skill and Assessment of Factors Affecting Test Analysis and Evaluation: The Case of Three Selected High Schools in Addis Ababa. Unpublished thesis. Retrieved on 18 August 2016 from <http://www.AAU.et>.
- MoE. (1994). *Education and Training Policy*. Addis Ababa: St. George Printing Press.
- MoE. (2002). *The Education and Training Policy and Its Implementation*. Retrieved on 20 August 2016 from <http://planipolis.iiep.unesco.org>
- MoE. (2011). *English for Ethiopia: Student Textbook Grade 7*. Uganda: MK Publishers Ltd.
- MoE. (2011). *English for Ethiopia: Teacher Guide Grade 7*. Uganda: MK Publishers Ltd.
- Morayo & Ohia, I.N. (2014). Teacher-Made Language Test Planning, Construction, Administration and Scoring in Secondary Schools in Ekiti State. *Journal of Education and Practice*. 5(18) PP. 71-75. Retrieved on 15/11/2016 from <http://www.iiste.org>
- Muijs, D. (2004). *Doing Quantitative Research in Education with SPSS*. London: Sage Publication.
- O'Leary, Z. (2004). *The Essential Guide to Doing Research*. London: Sage Publication .
- Richards, J. C. & Schmidt, R. (2002). *Longman Dictionary of Language Teaching and Applied Linguistics*. 3rd ed. London: Pearson Education.
- Siddiek, A. G. (2010). The Impact of Test Content Validity on Language Teaching and Learning. *Asian Social Science*. 6 (12). PP.133-143. Retrieved in December, 2010 from <http://www.ccsenet.org/ass>
- Singh, Y. K. (2006) *Fundamental of Research Methodology and Statistics*. New Delhi: New Age International.

- Teshome Soromessa. (2015). The Practice of Student Assessment: The Case of College of Natural Science, Addis Ababa University, Ethiopia. *Journal of Education and Practice* 6 (25). P 141-149. Retrieved January 2017 from <http://www.iiste.org>
- Tharenou, P., Donohue, R. & Cooper, B. (2007). *Management Research Methods*. Cambridge: Cambridge University Press.
- Ur, P. (1996). *A Course in Language: Practice and Theory*. Cambridge: CUP.
- Yalew Endawok. (2011). የምርምር መሰረታዊ መርሆች ና አተገባበር፡፡ [Applications of basic research principles.] 3rd ed. Addis Ababa: Commercial printing enterprise.
- Yildirim, O. (2010). Washback Effects of a High-Stakes University Entrance Exam: Effects of the English Section of the University Entrance Exam on Future English Language Teachers in Turkey. In P. Robertson & R. Nunn (Eds.), *The Asian EFL Journal Quarterly* Volume 12(2) PP. 92-116 <http://www.asian-efl-journal.com>.
- Zimmaro, D. M. (2004). Writing Good Multiple-Choice Exams. The division of instructional innovation and assessment. Retrieved on August 19, 2004 from <http://www.utexas.edu/academic/diia>.

Appendix A: Questionnaire to Be Filled by English Language Teachers

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF HUMANITIES AND LAW

DEPARTMENT OF ENGLISH LANGUAGE

(Post Graduate Program)

Questionnaire to Be Filled by English Language Teachers

Dear teachers,

The researcher is currently conducting a study as a partial fulfillment of the requirements for the degree of Masters of Art in TEFL. Hence, the main purpose of this questionnaire is to gather valuable information for the research entitled “**The Practice of English Language Test Items Construction and Administration in Second Cycle Primary Schools.**” The success of this study will highly depend on your trustful response and so I hope you will give accurate and honest response for the questions presented. I assure you that your response will be kept confidential and used only for this research purpose. I would like to express my heartfelt thanks for your willingness to fill in this questionnaire. Please, read the instruction before attempting to answer the questions and give only one answer to each item unless you are told to do otherwise.

Thank you in advance again!

General Directions

- Do not write your name on any part of the questionnaire.
- Please put a tick (✓) mark in the corresponding response in the table.
- Give brief response to the open ended items where necessary.
- Please don't leave the questions unanswered.

Part I. Demographic Information

Direction: Indicate your answer by putting a tick (✓) mark in the given box and also write the name of the school in the space provided.

1. Name of school _____

2. Sex Male ☐ Female ☐

3. Age group

21-25	26-30	31-35	36-40	41-45	46 and above
☐	☐	☐	☐	☐	☐

4. Academic Status

Diploma	Degree
☐	☐

5. Academic Field: English as a major field ☐

English as a minor field ☐

Stream (English, Afan Oromo and Amharic majors) ☐

English neither as a major nor a minor ☐

6. Teaching experience (in years)

1-5	6-10	11-15	16-20	21 and above
☐	☐	☐	☐	☐

Part II: Questionnaire

Direction: The following items in the tables below are accompanied with five options that show frequency. Please respond to these items putting a tick (✓) mark in the corresponding space as per codes:

- Always (4),
- Usually (3),
- Sometimes (2),
- Rarely (1) and
- Never (0).

S.N.	A. Practices and Guidelines in Test Item Construction	Frequency				
		4	3	2	1	0
1.	Planning Classroom Tests					
1.1	I define measurable instructional objectives that help me construct qualified test items.					
1.2	I categorize instructional objectives in the levels of domains of Bloom's taxonomy.					
1.3	I determine the objective(s) of my test before constructing the test.					
1.4	I prepare a table of specification for the construction of my test.					
1.5	I prepare a test draft before constructing my test.					
1.6	I determine the content area to be covered before constructing the tests.					
2.	Constructing Classroom Tests					
2.1	I base each item on an instructional objective of the content depending on what has been taught in class.					
2.2	I concentrate on the ability to use language for real-life communication.					
2.3	I determine the test item formats I want to give considering the levels such as age and grade of my students.					
2.4	I construct a test with different item formats (e.g., multiple choice, True-False, etc.) without copying from the textbook.					
2.5	I determine the length of a test myself considering the levels such as age and grade of my students.					
2.6	I determine the difficulty of the items considering the levels such as age and grade of my students.					
2.7	I avoid items that are easily guessed without proper studying.					
2.8	I use authentic language in my test.					
2.9	I construct tricky items.					
2.10	I avoid unintended clues such as grammatical, specific determiners, rote verbal association, length of answers, and length of blank, etc.					
2.11	I ensure the reliability of my test constructing test items clearly following the principle of writing qualified test items formats.					
2.12	I make sure the validity of my test using an adequate sampling of the content taught in class.					
2.13	I avoid using contents which reflect bias such as gender, racial, or etc.					
2.14	I write contextualized items.					
2.15	I use appropriate wording to the level of education where the item will be used.					
2.16	I ensure the readability of my tests by using clear language.					

S.N	Practices and Guidelines in Test Item Construction	Frequency				
		4	3	2	1	0
3.	Revising Classroom Test					
3.1	I arrange responses in a systematic order or patterns for the sake of scoring items easily.					
3.2	I ensure that each item is independent of all other items.					
3.3	I ensure that there is a single clearly correct answer for each item.					
3.4	I proofread my test to avoid errors such as grammatical, spelling and typing before reproducing it.					
3.5	I arrange test items from easy to difficult.					
3.6	I arrange test item formats in appropriate order.					
3.7	I give my test to other colleagues in order to comment it.					
3.8	I consider the comments and criticism offered by other colleagues in order to improve the quality of my test.					

Open Ended Questions

1. What does happen when test items in their formats are not constructed following their guiding principles?

2. What do you think of the practicality of test items construction guiding principles?

3. What are the possible solutions for the problems of the practicality of test items construction?

S.N.	B. Practices and Guidelines in Test Administration	Frequency				
		4	3	2	1	0
4.	Pre-test Administration					
4.1	I prepare my test long before its administration so that it is not done in a hurry.					
4.2	I announce a test at least a week before.					
4.3	I inform my students about the time of test administration.					
4.4	I check for the readiness of all materials for test administration.					
4.5	I remind students to get to the classroom early.					
4.6	I give my students practice tests.					
4.7	I identify the negative markers: underline, type them in capital letters or put them in boldfaced type.					
4.8	I inform the students what types of questions will be on the test (essay, multiple-choice, etc.) prior to the examination.					
4.9	I indicate how much the examination will count toward determining the final grade.					
4.10	I avoid constructing time-consuming tests.					
4.11	I write clear instructions.					
4.12	I keep test questions securely.					
4.13	I prepare enough copies of the test papers.					
4.14	I reproduce aesthetically pleasing copies.					
4.15	I administer too long tests that bring administration difficulties like test anxiety.					
5.	While Test Administration					
5.1	I prefer to administer my test myself.					
5.2	I distribute test papers properly.					
5.3	I remind my students to mentally figure out how much time they will need for each part.					
5.4	I inform students to look over the whole test quickly in order to get a good grasp of its different parts after the test is distributed.					
5.5	I administer my tests considering conducive environment such as light, sitting arrangement, weather condition, etc.					
5.6	I consider my students' psychological conditions.					
5.7	I make proper supervision while administering my test					
5.8	I handle any case of exam malpractice (e.g. cheating) in a regulated way.					
5.9	I advise my students to concentrate on the test as carefully as possible.					
5.10	I allow sufficient time for test completion.					
5.11	I inform students a few minutes before the end of test time so that they can finish on time, proofread their answers, and catch errors.					

S.N	B. Practices and Guidelines in Test Administration	Frequency				
		4	3	2	1	0
6.	Post-test Administration					
6.1	I collect answer sheets properly.					
6.2	I encourage my students to discuss to evaluate the quality of the test items immediately.					
6.3	I give immediate feedback to my students.					
6.4	I encourage my students to comment on the general test administration procedures.					

Open ended questions

1. List down the problems of test administration you encountered:

A. Before the test administration

B. While test administration

C. After test administration

2. What are the possible solutions for the problems you encountered above?

Appendix B: Questionnaire to Be Filled by Grade 7 Students

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF HUMANITIES AND LAW

DEPARTMENT OF ENGLISH LANGUAGE

(Post Graduate Program)

Questionnaire to Be Filled by grade 7 students

Dear students,

The researcher is currently conducting a study as a partial fulfillment of the requirements for the degree of Master of Arts in TEFL. Hence, the main purpose of this questionnaire is to gather valuable information for the research entitled “**The Practice of English Language Test Items Construction and Administration in Primary Second Cycle Schools.**” The success of this study will highly depend on your trustful response and so I hope you will give accurate and honest response for the questions presented. I assure you that your response will be kept confidential and used only for this research purpose. I would like to express my heartfelt thanks for your willingness to fill in this questionnaire. Please read the instruction before attempting to answer the questions and give only one answer to each item unless you are told to do otherwise.

Thank you in advance again!

General Directions

The Researcher

- Do not write your name on any part of the questionnaire.
- Please put a tick (✓) mark in the corresponding response in the table.
- Give brief response for the open ended items where necessary.
- Please don't leave the questions unanswered.

Part I. Demographic Information

Direction: Indicate your answer by putting a tick (✓) mark in the given box and also write the name of the school in the space provided.

7. Name of school _____

8. Sex Male ☐ Female ☐

9. Age group

11-15	16-20	20 and above
☐	☐	☐

Part II: Questionnaire

Direction: The following items in the tables below are accompanied with five options. Please respond to these items putting a tick (✓) mark in the corresponding space as per codes: **always (4), usually (3), sometimes (2), rarely (1) and never (0).**

S.N.	Practices and Guidelines in Test Administration	Frequency				
		4	3	2	1	0
1.	Pre-test Administration: Our English teacher					
1.1	announces a test at least a week before.					
1.2	informs us about the time of test administration.					
1.3	checks for the readiness of all materials for test administration.					
1.4	reminds us to get to the classroom early.					
1.5	gives us practice exercises before the test.					
1.6	identifies the negative markers; underline, type them in capital letters or put them in boldfaced type.					
1.7	informs us what types of questions will be on the test (essay, multiple-choice, etc.) prior to the examination.					
1.8	indicates to us how much the examination will count toward determining the final grade.					
1.9	arranges proper sitting arrangement.					
1.10	administers a test at the proper time of the day putting into consideration the weather condition and other factors.					
1.11	writes concise instructions.					
1.12	keeps test questions securely.					
1.13	prepares enough copies of the test papers.					
1.14	ensures that there is enough light in classroom.					
1.15	administers too long tests that bring administration difficulties like test anxiety.					
1.16	reproduces aesthetically pleasing copies.					
1.17	uses different types of test item formats.					
1.18	constructs tests free from grammatical errors.					
1.19	constructs tests free from spelling errors .					
1.20	constructs tricky items.					
2.	While Test Administration: Our English teacher					
2.1	distributes test papers properly.					
2.2	reminds us to mentally figure out how much time we will need for each part of the test paper.					
2.3	advises us to concentrate on the test as carefully as possible.					

S.N	B. Practices and Guidelines in Test Administration	Frequency				
		4	3	2	1	0
2.4	informs us a few minutes before the end of the test time so that we can finish on time, proofread our answers, and catch errors.					
2.5	prefers to administer his/her test him/herself.					
2.6	informs us to look over the whole test quickly in order to get a good grasp of its different parts after the test is distributed.					
2.7	makes proper supervision while administering his/her test.					
2.8	handles any case of exam malpractice (e.g. cheating) in a regulated way.					
2.9	allows us sufficient time for test completion.					
2.10	considers our psychological conditions.					
2.11	administers tests without teaching us the contents.					
3.	Post-test Administration: Our English teacher					
3.1	collects our answer sheets fairly.					
3.2	encourages us to discuss to evaluate the quality of the test items immediately.					
3.3	gives us immediate feedback.					
3.4	encourages to comment on the general test administration procedures.					

Open ended questions

3. List down the problems of test administration you encountered:

D. Before the test administration

E. While test administration

F. After test administration

4. What are the possible solutions for the problems you listed above in No 1?

Appendix C: Questionnaire to Be Filled by Grade 7 Students Amharic Version

ኢዳማ ሳይንስ እና ቴክኖሎጂ ዩኒቨርሲቲ

የሰነድና ህግ ት/ቤት

የእንግሊዝኛ ቋንቋ ት/ርት ክፍል

የድህረ-ምረቃ ፕሮግራም

በ7ኛ ክፍል ተማሪዎች የሚሞላ የፅሁፍ መጠይቅ

ዉድ ተማሪዎች፤

የዚህ ፕሮጀክት አድራጊ ለሁለተኛ ዲግሪያቸው ሚሚያ እንዲሆናቸው ጥናት በማካሄድ ላይ ይገኛሉ። የዚህ የፅሁፍ መጠይቅ ዋና ዓላማም የአንደኛ ደረጃ ሁለተኛ ሳይክል የእንግሊዝኛ ፈተና አዘገጃጀት እና አፈታተን ላይ ጥናት ለማካሄድ አስፈላጊ የሆኑ መረጃዎችን ለመሰባሰብ ብቻ ነው። ለዚህ ጥናት ስኬት የእናንተ ትክክለኛ እና ተአማኒነት ያለው መልስ ወሳኝ ነው። ስለሆነም ይህን እንደምትሰጡኝ እተማመናለሁ። የምትሰጡት መልስ ለዚህ ጥናት ብቻ እንደሚውል ላረጋግጥላችሁ እወዳለሁ። ለፈቃደኝነታችሁ ከልብ የመነጨ ምስጋናዬን አቀርባለሁ። መልስ ከመመለሳችሁ በፊት ትዕዛዙን በደንብ ያንብቡ። ለእያንዳንዱ ጥያቄ ሌላ ማብራሪያ ካላስፈለገው በስተቀር አንድ መልስ ብቻ ይስጡ።

በድጋሚ ከልብ አመሰግናለሁ!

ጠቅላላ ትዕዛዝ

ጥናት አድራጊዉ

- በማንኛውም መጠይቅ ላይ ስም አይፃፍም።
- ከፊት ለፊት ባለው ሰንጠረዥ ዉስጥ የ (✓) ምልክት ያስቀምጡ።
- ማብራሪያ ለሚፈልጉ ጥያቄዎች ማብራሪያዉን በተሰጠው ስፍራ ላይ ያስቀምጡ።
- መልስ ሳትሰጡ አትለፉ።

ክፍል አንድ: ግላዊ መረጃዎች

መመሪያ አንድ: መልስህ/ሽን በተሰጠው ሳፕን ዉስጥ የ (✓) ምልክት በመጠቀም መልስ/ሽን።

ሀ. የት/ቤት ስም:-----

ለ. ፆታ ወንድ ☐ ሴት ☐

ሐ. ዕድሜ

h11-15	h16-20	20 እና ከ20 በላይ
☐	☐	☐

ክፍል ሁለት: መጠይቅ

መመሪያ: በሚቀጥለው ሰንጠረዥ ዉስጥ ላሉት ጥያቄዎች ትይዩ 5 ምርጫዎች ተሰጥተዋል። ለጥያቄዎቹ ለአንተ/ቺ ተስማሚዉን መልስ የ(✓) ምልክት በማስቀመጥ መልስ/ሽን። ለቁጥሮቹ የተሰጡት መልሶች እንደሚከተለው ተገልፀዋል።

- | | |
|--------------------|---------------------|
| 0. በፍፁም አይተገብሩም | 3. አብዛኛውን ጊዜ ይተገብራሉ |
| 1. አልፎ አልፎ ይተገብራሉ | 4. ሁልጊዜ ይተገብራሉ |
| 2. አንዳንድ ጊዜ ይተገብራሉ | |

ተ.ቁ	የፈተና አሰጣጥ መመሪያ	የድግግሞሽ መጠን				
		4	3	2	1	0
1.	ከፈተና በፊት የሚከናወኑ ጉዳዮች፡ የእንግሊዝኛ መምህራችን					
1.1	ፈተና መኖሩን በትንሹ ከአንድ ሳምንት በፊት ያሳውቁናል፡፡					
1.2	የፈተናውን ሰዓት በአግባቡ ይነግሩናል፡፡					
1.3	ለፈተናው አስፈላጊ የሆኑትን ነገሮች ሁሉ ከፈተናው በፊት ያጠናቅቃሉ፡፡					
1.4	ወደ ፈተናው በጊዜ እንድንመጣ ያደርጉናል፡፡					
1.5	የሙከራ ፈተና ይሰጡናል፡፡					
1.6	በፈተናው ላይ የአሉታ አመልካቾች በግልፅ እንዲታዩ በማስመር ፣ በካፒታል ፊደላት ወይም አድምቀው ይፅፉናል፡፡					
1.7	በፈተናው ላይ የሚካተቱትን የጥያቄ ዓይነቶች ቀድመው ያሳውቁናል፡፡					
1.8	ፈተናው ምን ያህል እንደሚይዝ ቀድመው ያሳውቁናል፡፡					
1.9	የተማሪዎች መቀመጫ በአግባቡ የስተካከለ መሆኑን ያረጋግጣሉ፡፡					
1.10	ፈተና የሚፈትኑን የአየር ፀባዩን እና ሌሎች ሁኔታዎችን ከግምት ውስጥ በማስገባት ነው፡፡					
1.11	ፈተናው ላይ ግልፅ መመሪያ ይፅፉናል፡፡					
1.12	ፈተናውን በአግባቡ በሚሰጥ ይይዛሉ፡፡					
1.13	በቂ የሆነ የፈተና ኮፒ ያዘጋጃሉ፡፡					
1.14	በክፍል ውስጥ በቂ ብርሃን መኖሩን ያረጋግጣሉ፡፡					
1.15	በጣም ብዙ የሚያሰለች ጥያቄዎችን ይፈትኑናል፡፡					
1.16	ትራት ያለው የፈተና ኮፒ ይሰጡናል፡፡					
1.17	የተለያዩ የጥያቄ ዓይነቶች ይጠቀማሉ፡፡					
1.18	የሚያዘጋጁት ጥያቄ ከሰዋሰድ (Grammar) ችግር የፀዳ ነው፡፡					
1.19	የሚያዘጋጁት ጥያቄ ከፊደላት ስህተት (Spelling) የፀዳ ነው፡፡					
1.20	አሳሳች ጥያቄዎችን ያዘጋጃሉ፡፡					
2.	ፈተና በምንፈተንበት ወቅት፡ የእንግሊዝኛ መምህራችን					
2.1	ፈተናውን በአግባቡ ለተማሪዎች ያሰራጫሉ፡፡					
2.2	ፈተናውን ለመስራት ምን ያህል ጊዜ ለእያንዳንዱ ዓይነት ጥያቄ እንደሚያስፈልግ ያሳውቁናል፡፡					
2.3	የፈተናው ጊዜ ከማለቁ በፊት ምን ያህል ጊዜ እንደቀረ ይነግሩናል ከዚያም መልሶቻችንን እንደገና እንድናርም አመቺ ጊዜ ይኖረናል፡፡					
2.4	ፈተናውን በአግባቡ እንድንሰራ ያበረታቱናል፡፡					
2.5	ፈተናውን የሚፈትኑን መምህራችን ናቸው፡፡					
2.6	ፈተናው ሁሉንም ጥያቄዎች መያዙን ፈተና መስራት ከመጀመሪያ በፊት እንድናረጋግጥ ያደርጉናል፡፡					
2.7	ፈተናው በአግባቡ እየተሰጠ መሆኑን በትክክል ይቆጣጠራሉ፡፡					
2.8	ከፈተና ህግ ውጭ የሆኑ ጉዳዮችን በአግባቡ ይከታተላሉ፡፡					
2.9	ፈተናውን በአግባቡ እንድንሰራ በቂ ጊዜ ይሰጡናል፡፡					
2.10	የተማሪውን የስነ-ልቦና ሁኔታ ይጠብቃሉ፡፡					
2.11	ሳያስተምሩን ይፈትኑናል፡፡					

ተ.ቁ	የፈተና አሰጣጥ መመሪያ	የድግግሞሽ መጠን				
		4	3	2	1	0
3.	ከፈተናው በኋላ የእንግሊዝኛ መምህራችን					
3.1	የመልስ መስጫ ወረቀቱን በአግባቡ ይሰበስባሉ፡፡					
3.2	የፈተናውን ጥራት ወዲያው እንድንገመግም ያበረታቱናል፡፡					
3.3	መልሱን ወዲያው ያሳውቁናል፡፡					
3.4	በፈተናው አሰጣጥ ዙሪያ አስተያየት እንድንሰጥ ያደርጉናል፡፡					

ማብራሪያ የሚፈልጉ ጥያቄዎች

1. ከዚህ በፊት ፈተና ስትፈቱ ያጋጠማችሁን ችግሮች ፃፉ፡፡

1.1. ከፈተና በፊት

1.2. በፈተና ወቅት

1.3. ከፈተና በኋላ

2. ከላይ ለተጠቀሱት ችግሮች መፍትሄዉ ምንድን ነዉ ትላላችሁ?

Appendix D: Questionnaire to Be Filled by Grade 7 Students Afan Oromo Version

Yuunivarsiitii Saayinsii fi Teknolojii Adaamaa

Mana Barnoota Huumaanitiif fi Seeraa

Muummee Afaan Ingilizii

Sagantaa Guttannaa Eeb-booddee

Bargaaffii Barattoota Kutaa 7^{ffaa} guuttamu

Kabajamtoota Barattoota;

Kaayyoon guddaan bargaaffii kanaa haala qophii fi kenniinsa qormaata Ingiliffaa manneen barnoota sad.1^{ffaa} Saayikilii 2ffaa irrattii qorannoo adeemsisuuf odeeffannoowwan barbaachisaa ta'aan sassaabuuf dha. Galma gahinsa qorannoo kanaaf gaaffiilee dhihaataniif deebiin keessan sirrii fi amanamaa ta'uun murteessa dha. Kanaafuu deebii amanamaa ta'e akka naaf kennitan nan amana. Deebii isiin keennitaniif durseen isiin galateeffadha. Deebii deebisuu keessaniin dura qajeelfama sirriitti duubbisaa. Gaafiileen ibsaa dabalataa barbaadan yoo hin jirannee gaaffii tokkoof deebii tokko qofa kennitu.

Galatoomaa!

Qo'atticha

Ajaja Waliigalaa

- Bargaaffii kamiratiyyuu maqaan hin barraa'u.
- Unkaa gucarraa jiru irratti mallattoo (✓) ka'i.
- Gaaffilee banaa kennamaniif iddoo isiniif kennamerratti ibsa ifa ta'e kenni.
- Gaaffilee osoo hindeebisin irra hin utaalinaa.

Kutaa Tokko: Odeeffannoo Dhuunfaa

Ajaja Tokko: Deebii kee iddoo siif kennametti mallattoo (✓) fayyadamuun deebisi.

A. Maqaa Mana barumsa _____

B. Saala Dhi ☐ Dub ☐

C. Umrii

11-15	16-20	20 and above

Bargaafilee

Ajaja II. Gaaffilee armaan gaditti guca keessatti mul`atan filannoowwan shan shan qaban. Gaaffilee dhiyaatan fuulduratti filannoowwan kennaman keessaa kan deebii ta`an jettanii yaadanirratti mallattoo (✓) kaa`aa. Hikni lakkofsoota akka armaan gaditti dhihaateera.

- | | |
|-----------------------------------|-----------------------------------|
| 0. Kongumaa hin raawwatani. | 3. Yeroo irra caalaa ni raawwatu. |
| 1. Darbee darbee ni raawwatu. | 4. Yeroo mara ni raawwatu. |
| 2. Yeroo tokko tokko ni raawwatu. | |

T.L	Qajelfama Haala Kenniinsa Qormaataa.	Haala Raawwii				
		4	3	2	1	0
1.	Qormaata Dura Wantoota Raawwataman: Barsiisaan Afaan Ingiliizii Keenya:					
1.1	Qormaatan jiraachuu yoo xiqqaate turban dura nutti hima.					
1.2	Yeroo(Sa`aatii) qormaataa nutti hima.					
1.3	Waan barbaachisu hundaa qormaata dura xumura.					
1.4	Yeroon gara qormaataa akka dhufnu nitaasisu.					
1.5	Qormaata yaalii nuuf kennu.					
1.6	Qormaataratti wantoota ifa hin taane akka mul`atutti nutti agarsiisu.					
1.7	Gaaffilee qormaataa kanneen akkamii akka ta`an nutti himu.					
1.8	Qabbiyyee qormaatichaa hangam keessaa akka qabamu nu beeksisu.					
1.9	Teessoon barattootaa seeraan taa`a.					
1.10	Qormaata kan nuqoran haala qilleensaa fi wantoota biro ilaalcha keessa galchuudhani.					
1.11	Qormaaticharratti qajeelfama ifa ta`e barreessu.					
1.12	Qormaaticha seeraan iccitiidhan qabu.					
1.13	Baay`ina qormaataa gahaa qopheessu.					
1.14	Daree qormaataa keessa ifni gahaan jiraachuu nimirkaneeessu.					
1.15	Gaaffilee baay`ee nuffisiisoo <b>Fakkeenyaaf</b> Gaaffilee baay`ee fi dheera, ta`an nuqoru.					
1.16	Qormaanni isaan nuuf kennan barreeffamni isaa qulqulluudha.					
1.17	Gosoota gaafii gara garaa ni fayyadamu. Fakkeenyaaf Dhugaa soba, Firoomsi, Filadhu, Deebii sirri kenni/guutii fi Gaafiilee barreeffama.					
1.18	Gaaffileen isaan qopheessan rakkoo caaslugatirraa (Grammar) bilisa.					
1.19	Gaaffileen isaan qopheessan dogoggora qubeetirraa (Spelling) bilisa.					
1.20	Gaaffilee nama dogoggorsan ni dhiyeessu.					
2.	Wayta Qormaata Qoramnu: Barsiisaan Afaan Ingiliizii Keenya:					
2.1	Qormaaticha barattootaaf seeraan ni raabsuu.					
2.2	Tokkoon tokkoon gaaffilee hojjachuuf yeroo hangamii akka barbaachisu nu beeksisu.					
2.3	Qormaaticha sirritti akka hojjannu nu jajjabeessu.					

	Qajelfama Haala Kenniinsa Qormaataa	Haala Raawwii				
		4	3	2	1	0
2.	Wayta Qormaata Qoramnu: Barsiisaan Afaan Ingiliizii Keenya:					
2.4	Yeroon qormaaticha itti hojjatamu xumuraamuun dura yeroon hammamii akka hafe nutti himu. Boodas deebin keenya sirrii ta`uu irra deebinee akka ilaallanuuf yeroo gahaa arganna.					
2.5	Qormaaticha ofumaaf nuqormaatu.					
2.6	Qormaatichi gosoota gaaffilee hundumaa hammachuu isaa akka mirkanneeffanu nu taasisu.					
2.7	Qormaatichi seeraan keennamaa jiraachuu isaa ni hordofu.					
2.8	Wantoota seera qormaataatiin ala ta`an sirritti hordofu.					
2.9	Yeroo gahaa ta`e nuuf kennu.					
2.10	Xiinsammuu barattootaa nieegu.					
2.11	Osoo nu hin barsiisin qormaata nuuf kennu.					
3.	Qormaata booda barsiisaan: Afaan ingiliizii keenya:					
3.1	Waraqaa deebii seeraan barattootarraa fuudhu.					
3.2	Qulqullina qormaataa akka gamaggamnu nu jajjabeessu.					
3.3	Deebii yoosuma nuuf deebisu					
3.4	Haala qormaanni ittii kennameerratti yaada akka kenninu ni taasisu.					

Gaaffilee Banaa

1. Rakkoolee yemmuu qormaata qoramtu si muudatan tarreessi.

A. Qormaaticha qoramuun dura

B. Yemmuu qormaaticha qoramtu

C. Erga qormaaticha qoramteen booda.

2. Rakkoolee armaan olitti lakkoofsa (1) jalatti ka`aniif furmaanni isaa maal jetteet yaadda?

Appendix E: Classroom Observation Checklist on Test Administration

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF HUMANITIES AND LAW

DEPARTMENT OF ENGLISH LANGUAGE

Classroom Observation Checklist on Test Administration

The main purpose of this observation checklist is to gather information observing the activities and situations happening in the classroom concerning the administration of English language tests. The activities will be marked in the corresponding categories depending on the occurrence. Hence, put a tick (✓) mark corresponding the given responses in the following table.

Part I. General Information

Name of the School: _____ Date: _____ Time begin: _____

Observed Classroom: _____ Time ended: _____

Part II: Observation Checklist on the administration of English Language tests.

S.No	Principles of test administration	Yes	Uncertain	No
1	Actions Before Test Administration			
1.1	All materials are ready before the administration of the test.			
1.2	The weight of the test is determined.			
1.3	The length of the test is manageable for the students.			
1.4	The instructions are absolutely clear.			
1.5	There are enough copies of the test papers.			
1.6	The copies are aesthetically pleasing.			
1.7	There are different types of test item formats on the test.			
1.8	There are enough desks in class.			

S.No	Principles of test administration	Yes	Uncertain	No
1.	Actions in Pre-test administration			
1.9	The desks arranged at distance apart.			
1.10	There is enough light in classroom.			
1.11	All students were on time before the administration of the test.			
1.12	Test papers are kept properly.			
2	Actions While Test Administration			
2.1	English teachers administer their own test.			
2.2	There is fair distribution of test papers.			
2.3	English teachers handle any case of exam malpractice (e.g. cheating) in the regulated way.			
2.4	The teachers keep the psychological conditions of students.			
2.5	Teachers advise students to concentrate on the test.			
2.6	Enough time is given for the students.			
2.7	The test is administered at the proper time of the day.			
2.8	English teachers give corrections within the appropriate time.			
2.9	Teachers announce time periodically.			
2.10	Teachers control cheating effectively.			
3	Actions After Test Administration			
3.1	Students have completed the test within the given time.			
3.2	Answer sheets are collected fairly.			

Additional Remarks

Appendix F: English language Test items construction Evaluation Checklist

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF HUMANITIES AND LAW

DEPARTMENT OF ENGLISH LANGUAGE

English language Test items construction Evaluation Checklist

The main purpose of this checklist is to gather concrete information concerning the construction of English language test items. The items will be evaluated based on the corresponding guidelines. Hence, put a tick (✓) mark under the corresponding response.

Part I. General Information

Name of the School _____ Academic year _____ Semester _____

Part II: English Language Test Items Construction Checklist

S.N	Test Items Evaluation	Yes	Uncertain	No
1.	Multiple Choice Items			
1.1	Are the multiple choice items contextualized?			
1.2	Is the question in the stem definite?			
1.3	Are negatively stated stems used sparingly?			
1.4	Is the negative word easily identified e.g. underlined or capitalized?			
1.5	Are the correct responses randomly distributed?			
1.6	Are the unintentional clues (grammatical, phonological, morphological, and so forth) avoided?			
1.7	Are the distractors plausible?			
1.8	Are the alternatives in an item the same in length or in a similar format (i.e., all phrases, all sentences, etc.)?			
1.9	Are the alternatives “none of the above” and “all of the above” used sparingly?			
1.10	Are capital letters (A, B, C, and D) used as response signs to avoid confusion?			
1.11	Are the alternatives listed on separate lines vertically?			

S.N	Items	Yes	Uncertain	No
2.	True-False Items			
2.1	Are the true-false items absolutely true or false?			
2.2	Are the items expressed in simple and clear language with a single idea in each test item?			
2.3	Are the statements worded appropriately?			
2.4	Are the sequences of true and false statements randomized to avoid discernible patterns of answers?			
2.5	Are the true and false items made of approximately equal average length throughout the test?			
2.6	Are the absolute clues (modifiers) like all, always, absolutely, never, rarely, most often, and so forth avoided?			
3.	Matching items			
3.1	Is the direction which is the basis for matching the stimuli with the responses clearly stated?			
3.2	Are the matching items homogeneous?			
3.3	Are all matching items on the same page?			
3.4	Are the two columns equal in number?			
4.	Short answer items(completion and short answer type)			
4.1	Are only significant words omitted from the statement?			
4.2	Are there any obvious and grammatical clues to the correct response?			
4.3	Is there only one concise correct response?			
4.4	Is there explicit direction for clarity?			
4.5	Are the blank spaces put at the end of the statement?			
4.6	Is the length of the blank space the same throughout the statements?			
5.	Essay Items			
5.1	Is the item defined clearly?			
5.2	Is there clear instruction for essay items?			
5.3	Are there optional questions on an essay test?			
5.4	Is there appropriate amount of time given for each question?			
5.5	Are the length and format of the response determined?			
6.	Overall Evaluation of test items construction			
6.1	Is each item fair for the level of education?			
6.2	Are the items expressed in the simple language?			

S. N	Items	Yes	Uncertain	No
6.	Overall Evaluation of test items construction			
6.3	Have bias such as race, gender, and nationality been avoided?			
6.4	Is there relevant information presented?			
6.5	Is there a single clearly correct (or best) answer for each item?			
6.6	Are the items concentrated on the ability of the students to use language for real-life communication?			
6.7	Are the items easily guessed without proper studying?			
6.8	Are most test item formats included?			
6.9	Is the length of the test fair for the class level?			
6.10	Is the language authentic?			
6.11	Are there any tricky items?			
6.12	Are the items independent of all other items?			
6.13	Are the items copied from the textbook?			
6.14	Are the items free from errors such as grammatical, spelling and typing?			
6.15	Are the items and their options spaced clearly on the same page?			
6.16	Are the items arranged from easy to difficult?			
6.17	Are the test item formats arranged properly?			
6.18	Is the test valid in containing adequate sampling of contents?			

Additional Remarks

Appendix G: Interview Questions for Language Department Heads

ADAMA SCIENCE AND TECHNOLOGY UNIVERSITY

SCHOOL OF HUMANITIES AND LAW

DEPARTMENT OF ENGLISH LANGUAGE

(Post Graduate Program)

Interview Questions for Language Department Heads

1. Do you believe that English language teachers in your CRC schools have enough knowledge in constructing English language test items and administering the items well? How do they construct and administer English language test items?
2. Do you think that English language teachers in your CRC schools usually apply the principles of test items construction and administration (e.g. using table of specifications, objectives, test item formats etc.)? What is your evidence?
3. Have you got English language teachers who did not construct and administer tests according to the principles? If so what is the reason behind?
4. Is there any trend of evaluating test items (e.g. giving to your colleagues) before they are reproduced and administered to the students? If so what is your base?
5. What about the trend of evaluating test items (e.g. with colleagues or/and students) after they are administered to the students? If so what is your base?
6. What problems did you observe frequently from the result of your evaluation?
7. What do you suggest on the practical application of the principles of test items construction and administration?

Appendix H: English Language Test Paper in the First Semester 2016/17

In _____ CRC English final exam.1 st Sem .for Grade 7 th.in 2009E.C.

I. Write Ttrue " if statement is correct and" False" if statement is not correct.

1. we use" many " with countable nouns.
2. And But and becousen are conjuccion .
- 3,we use much with un countable nouns.
4. comma is used to separate items on a list.
5. A question mark is used after a questions.

II. Match with their meanings of the words

"A"

"B"

- | | |
|----------------|----------------------------------|
| 6. Demoralised | A. war |
| 7.Colony | B. old enough for harvesting |
| 8. Advance | C.well known (popular) |
| 9. Famous | D. move for ward |
| 10. mature | E. Country controlled by another |
| 11. Battle | F. last confidance and hope |

III. Choose the best answer from the given choice

12. _____he has many brothers A. is B. Does C. Did D. Do
13. He can speak Somalia _____he can't speak walayta.
A. But B/ Both C. Because D. can't
14. You should walk _____ whe crossingthe road .
A. corelely B. loudly C. Silently D. carefully
15. She _____to church every Sunday .A. goes B. go C. going D. does
16. He will come _____ Sunday . A. at B. on C. in D. to
17. Where _____you live. A. is B. am C/ do D. be
- 18 . I _____ very well what about yoyu? A. are B. am C. is D.good
19. Neither Ali _____Bontu went to Asella. A. so B/ nor C. about D. none
20. Last year we _____maize . A. harvest B. havests C. harvested D. harvesting
21. Which one is different from the athor ? A. always B. after C. some times D. used to
22. As they _____,it started raining
A. where fighting B. was fighting C. is fighting D. a and b
23. Do you sell fish? A. yes I do B. No ,he do C. Yes ,he is D. No ,I do
24. Our class didn't visit Lalibela. A. Neither I am B. Neither did I C. Neither do I D. all
25. _____this the way to lalibela .A. are B. wher C. is D. have.

Set by Language Dipp

Appendix I: English Language Test Paper in the First Semester 2016/17

2009

CRC First Semester English Exam For Grade 7

Write 'True If The Statement Is Correct And "False " If Not

1. But is used to join two opposite ideas
2. The meaning of famous is known.
3. Conjunctions are words used to join two words or phrases .
4. "Wh" questions need answer.
5. The past tense of "receive " is "received"

II. Match column "A" with their responses from "B"

A

B

- | | |
|--------------|----------------------|
| 6. Why | A. is your town? |
| 7. How big | B. coat is it? |
| 9. What | D. are you laughing? |
| 8. Whose | C. is her new car? |
| 10. How fast | E. time is it ? |

III. Choose The Best Answer From Following Question

11. Gudeta _____ not play net ball.
a. do b. does c. have d. has
12. He can speak shinasha _____ the can not write it;
a. and b. but c. or d. so
13. I couldn't answer the question _____ I don't know its meaning.
a. because b. so c. and d. but
14. last year our teacher _____ us science.
a. Teachers b. teach c. teaching d. taught
15. _____ do you born in Addis Ababa or at Adama?
a. what b. which c. when d. where
16. Is there any water in the bottle? No there is _____ in it
a. no b. any c. some d. little
17. She _____ at Adama. a. live b. lives c. lived d. living
18. What is your mother doing? she is _____
a. cook b. cooking c. cooked d. cooks
19. While we _____ TV, my sister was studying.
a. watched b. are watching c. was watching
d. were watching
20. Zara studied hard and _____ Ash .
a. so do b. so did c. so can d. could
21. I like oranges _____ it is sweet.
a. but b. or c. because d. as a result
22. Merga is going to Gonder .
a. so am I B. Neither am I c. either am I d. A&B
23. I _____ tried on yeast ter day a. is b. was c. are d. were
24. Who is _____ student in the class?
a. the tau b. the tallest c. the toller d. the most tallest
25. Can you read and speak? a. yes I can b. yes I can't c. no I'm d. No I can
26. Neither the director _____ the teacher come on time .
a. and b. either c. nor d. all
27. _____ subject do you like English or Afan Oromo?
a. Where b. which c. when d. who
28. Ahmed _____ Sofia can speak Arabic
a. or b. and c. but d. all
29. They usually _____ their home during holy days.
A. visit b. decorate c. celebrate d. tour
30. He is a lazy student _____ he passed the examination.
a. and b. or c. but d. so

Have good chance by language dip. t

Appendix J: English Language Test Paper in the First Semester 2016/17

In CRC In 2009 E.C 1st semester English Exam For Grade 7th

I. Write True For Correct Statement And Write False For Incorrect Statement

1. Superlative degree is used to compare 3 or more things.
2. Either means one of the two neither means none of the two.
3. Since is used with a point of time for used with length of time.
4. Prepositions are words or groups of words.
5. Went & gone are the past & past participles form of go.
6. A and an are indefinite articles.
7. Letter can be classified into 4 categories.
8. Capitalization is not punctuation.
9. Continent is a large mass of land.

II. Match the following nouns adjective & adverbs with their kinds and write question words.

- | A | B |
|---|-------|
| 10. There are ----- kinds of nouns | A, 5 |
| 11. There are ----- kinds of adjectives | B, 6 |
| 12. There are ----- kinds of adverbs | C, 7 |
| 13. Who, where, when, why ----- | D, WH |

III. Choose The Best Answer From The Given Alternatives.

14. May is used to ----- A, express possibility /doubt/ B, express permission C, express wish/hoped/ all are correct
15. Can is used to ----- A, express action taking place at the time of speaking D, all are correct
16. Which one is parts of letters? A, your address and the date B, reference & sign C, introduction D, main part of conclusion E, all
17. Africa is one of the ----- continent of the world. A, 7 B, 56 C, 9 D, 4
18. ----- is the largest continent in the world. A, Asia B, Africa C, Nigeria D, Egypt & Cairo
19. ----- is capital city of Egypt. A, Cairo B, Tripoli C, Algier D, Kampala
20. ----- is the holy month of Muslims. A, Muharrar B, Ramadan C, Fasika D, Genna
21. Shall and will indicate -----? A, promise B, future time C, intention D, command determination E, all are correct
22. Abanesh is going to Gonder. A, so am I B, either am I C, neither am I D, so do I
23. While I was ----- at my aunt's home the war broke out. A, staying B, run C, watch D, visiting E, all are correct
24. Which one is skill of language? A, speaking B, reading C, writing D, all
25. Mohammed ----- many letters to his friends every day. A, wrote B, writes C, writing D, written.

IV. Complete /Fill/The Blank Space

26. The battle of Adwa took place in Adwa on 1st march -----
27. Ethiopian coffee is very -----
28. Delicious food prepared by the cook this statement is ----- voice.
29. Hailu eats injera for his lunch ----- voice past form of. A, do ----- B, is ----- C, go -----
30. Write 4 major language division in Ethiopia

- A, -----
- B, -----
- C, -----
- D, -----

Appendix K: English Language Test Paper in the First Semester 2016/17

CRC First semester English Final examination for Grade 7 in 2009 E C

Name _____ Grade _____ No _____

I. Write True if the statement is correct and False if it is Wrong.

- _____ 1. Both full stop and Question marks are the end marks.
- _____ 2. Formal letters are more personal than informal letters.
- _____ 3. the focus of a active voice is on the doer of the action.
- _____ 4. There is no any relation ship between origins of names and thir meaning.
- _____ 5. The beginning and Ending of a letter depends on who is to write it.

II. Match the meaning of words from coloum 'B' to 'A'

A

- _____ 6. Move up
- _____ 7. Four a place
- _____ 8. At this time
- _____ 9. Fire a gun
- _____ 10. Riler

B

- A. Visit
- B. Noun
- C. Climb
- D. Shoot
- E. King

III. Match with their ooposites from 'B' to 'A'

A

- _____ 11. Accept
- _____ 12. Hide
- _____ 13. Weap
- _____ 14. High
- _____ 15. Remember

B

- A. Show
- B. Laugh
- C. Refuse
- D. Low
- E. Forget

IV. Choose the best answer

- _____ 16. It was raining when we _____ out.
A. Go B. going C. gone D. Went
- _____ 17. I Don't like hot weather, but Aster _____
A. Does B. doesn't C. do D. all
- _____ 18. A comma is used to _____
A. Ask question B. Separate items C. Omission of letters D. in abbreviation
- _____ 19. A few means A. enuou gh amount B. almost none C. Some D. all
- _____ 20. My friend _____ for me whene I aeeived A. Waited B. Has waited C. Was waiting D. has been waiting
- _____ 21. She often _____ coffee. A. Drank B. Drunk C. drinks D. Drinking
- _____ 22. It is used to join similar ideas A. because B. but C. Although D. And
- _____ 23. Last year she _____ to Sof Umer for atour. A. went B. goes C. go D. going
- _____ 24. While we _____ Adwa, we met a world war veteran.
A. Was visiting B. Visited C. Were visiting D. Visit
- _____ 25. Which one is good idea. A. People shouldn't go to school. B. We should play the whole day. C. All school should have computers. D. All
- _____ 26. Sofia usually _____ to the Foot ball field
A. Hurry B. Hurries C. Hurrys D. Hurrid
- V. change the following active voice to passive voice**
27. we make butter from milk.
28. some body cleans these rooms every day.
29. people never invite me to parties.
30. Bortola speaks English ever day.

Appendix L: English Language Test Paper in the First Semester 2016/17

English Final Examination Grade 7 1st Semester in 2009 for CRC

I. Say True if statement is True, Say False if statement False

1. 'so' used to in order to do or intension of doing.
2. Past simple tense express action in the future.
3. Because express reason.
4. We use much with uncountable nouns.
5. Neither expresses negative idea.
6. Although is a conjunction.

II. Match word A with their meaning in B

- | <u>A</u> | <u>B</u> |
|-------------------------|---------------------------------|
| 7. Why | A. Is this new bus? |
| 8. Punctuation | B. Express reason |
| 9. The use of comma (,) | C. How, When, and Where |
| 10. Question mark (?) | D. Full stop |
| 11. Active voice | E. A letter was written by her. |
| 12. Passive voice | F. She wrote letter. |

III. Choose the correct answer from the following Questions

13. Have you _____ your homework?
A. do B. does C. done D. did
14. Neither Abdi _____ Abaru went to Bolo.
A. or B. no C. nor
15. _____ Ali or Roba went to Arboye.
A. Neither B. nor C. either D. Both
16. _____ Rabira and Sifan are speas English well.
A. Both B. are C. is D. none
17. He _____ TV whe his father is home.
A. watch B. watching C. watched D. All
18. My father _____ to office every morning
A. go B. goes C. do D. does
19. _____ she have many brothers?
A. do B. does C. don't D. All
20. _____ you grade seven students?
A. Was B. were C. Are D. is

Appendix M: English Language Test Papers in the Second Semester 2015/16

Grade 7 2008 2nd semester English final exam.

Name _____ Grade _____ No. _____

Part one match 'A' with 'B' with their answer to complet the praises.

A

- ____ 1. We are good ____
- ____ 2. He is punished for ____
- ____ 3. She doesn't go to school every day ____?
- ____ 4. He is a student ____
- ____ 5. ount is for a woman uncle is for ____
- ____ 6. The person who sells meat is ____

B

1. a man
2. does she
3. did
4. atwritting
5. So am I
6. Stealing
7. Bucher
8. So I am

Part two:- choose the best anser from the following four alfer natives

- ____ 7. Ethiopia ____ a green red and yellow flag.
1. has 2. have 3. had 4. had to
- ____ 8. The book of the students ____ very old.
1. are 2. is 3. has 4. have
- ____ 9. Ridng a horse is easier than ____ amule.
1. riding 2. foriede 3. rice 4. rode
- ____ 10. Tewabe ____ examination in Bira school ncw.
1. takes 2. taken 3. took 4. is taking
- ____ 11. Do you feel ____ now than you did yesterday?
1. worse 2. worst 3. badder 4. best
- ____ 12. Prim minister Hailemariam Desalegn is ____ democratic ____ the
Ethiopian people respected him.
1. so/that 2. so, a/that 3. such/that 4. too/that
- ____ 13. The man is short ____ he can reach the roof.
1. though 2. but 3. although 4. all
- ____ 14. She often goes to Adama ____ herparents live there.
1. but 2. and 3. although 4. because
- ____ 15. Who wishes ____ be a pilot.
1. that 2. on 3. to 4. can
- ____ 16. Muy nother always ____ to church every Sunday.
A. is going to 2. has gone 3. study 4. goes
- ____ 17. They have ____ just now.
1. escape 2. toes cape 3. escaping 4. escaped

- ___18. My friend doesn't like to dance. ____
 1. nor do I 2. or do I 3. so am I 4. so do I am
- ___19. Either the king or the soldiers ___ to risk at this time.
 1. was 2. are c. were 4. is
- ___20. I will never forget the day ___ I met you.
 1. which 2. whose 3. when 4. whom
- ___21. I don't like him very much. She ___ always complaining him
 1. is 2. has 3. does 4. am
- ___22. She ___ get up early when she was working in the factory.
 1. used to 2. issued to 3. was used 4. using
- ___23. My girl friend is as ___ as Tigist.
 1. most beautiful 2. more beautiful 3. the most beautiful 4. beautiful
- ___24. I usually goes to work _____ bus.
 1. in 2. on 3. by 4. for

Part three choose the correct word that best complet the paragraph.

When I was a child I used to ___25___ with mud. My mother used to
 ___26___ me not to do so. But I used to ___27___ them. I didn't used
 to ___28___ all their ___29___ at all. Their ___30___ used to make me
 angry.

- | | | | |
|-------------------|------------|-----------|-----------|
| ___25. 1. advice | 2. make | 3. play | 4. dance |
| ___26. 1. advise | 2. advice | 3. refuse | 4. accept |
| ___27. 1. accept | 2. make | 3. refuse | 4. regret |
| ___28. 1. satisfy | 2. regret | 3. give | 4. accept |
| ___29. 1. advice | 2. advise | 3. help | 4. work |
| ___30. 1. advice | 2. nagging | 3. advise | 4. all |

Sety By

Appendix N: English Language Test Papers in the Second Semester 2015/16

At the CRC of _____ English Final Exam for Grade 7th 2nd Sem. 2008

Name _____

I. Write True if the statement is correct write false if the statement is incorrect

1. Irregular and regular adjectives have the same rules.
2. We use do/does in present simple questions.
3. Children miss opportunities in life when they get married too early.
4. Should and must are modal verbs.
5. The word where refers to time.

II. Match the following words with their meaning

- | A | B |
|--------------------|--|
| 6. How much | A. Intention or plan |
| 7. Graze | B. Once duty to do something |
| 8. Achieve | C. Animals that kills and eats other animals |
| 9. Intelligent | D. Uncountable noun |
| 10. Responsibility | E. Feed on vegetations |
| 11. Predator | F. Get what one wants |
| 12. Dream | G. Very wise and clever |
| | H. Countable nouns |

III. Choose the best answer for the following question

13. We were really surprised _____ he arrived un expectedly. A. if B. When C. as soon as D. none of the above
14. I am really hungry. Let's go for dinner _____ the film finishes. A. before B. as soon as C. when D. A and B
15. They _____ going to teach families about HIV/AIDS. A. am B. is C. are D. all are correct
16. _____ out of school can lead to missing good chance in life? A. dropping B. cases C. attending D. none
17. I am very busy, but I will go shopping _____ I have the time. A. until B. when C. if D. B and C
18. Neither the students nor the teachers came to school yesterday because of the holiday, this means. _____ A. the students came but not teachers B. the teacher's came but not students C. both the teachers and the students were absent D. all are correct
19. From the following which one is an adjective? A. big B. dangerous C. slowly D. A and B
20. Gobena plays football _____ than kamila. A. good B. better C. best D. little
21. How _____ eggs do you needs? A. many B. much C. more D. most
22. Hala likes chocolate so do se Tola. This means _____ A. Chala likes checolate but not Tola B. both Chala and Tola like Checolate C. A&B D. None

IV. Fill in the blank spaces by the given word

- Doctor Family Tailor Driver
23. _____ is a person who produce crops
 24. _____ is a person who helps patients
 25. _____ is a person who saw cloth

Set by English Dep^t

Appendix O: English Language Test Papers in the Second Semester 2015/16

Name _____ Section 7 _____ No _____

CRC English Final Examinations for grade 7th 2nd Sem. 2008

I. Say true if the statement is True and False if not true

1. May and might are used to express a possibility that something is true.
2. Adjectives are words that describe noun or pronoun.
3. Prepositions are used to link one part of sentence with other.
4. Synonyms are words that have opposite meaning.
5. The subject and verb must fit grammatically.

II. Select the correct word or phrases that compute the dialogue

6. Azeb: How are you
Aster: _____ A. I'm fine B. How many C. How often D. How much
7. Asrat: _____ do you go to Cinema?
Alem: twice a week A. How much B. How often C. I'm happy D. thank you
8. Hailu: Nice to meet you.
Roman: _____ A. Nice too B. Nice to me you C. Nice to meet you too D. not at all
9. Aman: Do you eat you breakfast?
Soret: _____ A. Yes I did B. No didn't C. Yes, I do D. Yes, she is
10. Tomas: How old you will be?
Bekele: I _____ be fifteen A. am B. in C. will D. have

III. Choose the best answer the following questions

11. I don't like things that are _____ sweet. A. too B. to C. two D. for
12. We have _____ Composition about Adwa. A. write B. writing C. written D. wrote
13. While I _____ to school, I _____ down. A. was coming/fell B. is coming/fall
C. were coming/fall D. are/coming
14. Abebe _____ to Bonga last year. A. go B. gone C. went D. goes
15. I know I _____ speak and write Somali very well. A. can B. could C. can't D. would
16. Azeb went to forest _____ she saw different type birds. A. which B. who C. where D. but
17. Did you visited Gambella _____ Hawasa. A. or B. but C. to D. also
18. If my sister reads hard she _____ pass the exam. A. would B. will C. could D. can
19. Tola _____ to school every day. A. go B. goes C. gone D. went
20. She married a few years _____ A. latter B. ago C. last D. also
21. Rift valley is known for _____ scenic beauty A. for B. to C. by D. from
22. The test is _____ than the previous one. A. strong/heavy B. easier C. big D. highest
23. They _____ played foot ball since morning. A. have B. has C. had D. is
24. We shall **tour** national park. What does tour mean? A. visit B. sail C. refuse D. invite
25. The fish that she _____ was tastey. A. painted B. cooked C. dug D. touch

IV. Write the past simple and past participle form of these words

26. Think _____
27. Collect _____
28. Take _____
29. Give _____
30. Break _____

Set by language teacher s of CRC

Appendix P: English Language Test Papers in the Second Semester 2015/16

In CRC English Final Examination for grade 7th 2nd Sem. In 2008

Say true for correct sentence & say false for wrong sentence

1. Girls shouldn't get pregnant until they are married.
2. We use but to join the statement which have similar idea.
3. Every one should discourage early marriage.
4. A person who can not hear is known as deaf.
5. Most languages are spoken in Ethiopia.
6. Gebeta is not a kind of traditional game.
7. The government should protect the endemic animal.
8. We use teff flour to make enjera. This sentence is passive voice.

II. Join column "A" with column "B"

- | A | B |
|-------------------|--|
| 9. Active voice | A. A person who can't speak |
| 10. Passive voice | B. Two |
| 11. Both | C. The opposite of dirty |
| 12. Dump | D. Halima bakes bread at the bakery |
| 13. Clean | E. Rice is packed in sacks |
| 14. Were | F. The past form of fight |
| 15. Fought | G. Words like stand up, look at the roof |
| 16. Imperatives | H. The past form of are |

III. Choose the answers for the following questions

17. 'It English lesson think they are very ____ A. Dull B. Impossible C. Harmless D. Interesting
18. She speaks ____ Somali and English. A. and B. both C. can D. can't
19. How did you came to school today? A. slowly B. Sadly C. Quickly D. A and C
20. The test we are today was ____ than previous one. A. difficult B. more difficult
C. the most difficult D. none
21. Coffee beans are died ____ they are grown into fine coffee. A. to B. so that C. but
D. although
22. My head is paining me, I have ____ A. headache B. toothache C. Stomachache D. all
23. Medicine ____ be kept out of reach the children. A. ought to B. had better C. should
D. Shouldn't
24. ____ boys and girls helped by keeping the crops from wild animals. A. Both B. non
C. each D. a lot
25. ____ you eat a balanced diet, you will not be healthy. A. If B. because C. unless D. while
26. They are not strong ____ to win the game. A. too B. enough C. not enough
D. non of the above

IV. Write short answer for the given questions

27. The past form of go _____
28. The past form of buy _____
29. Your fathers mother are called _____ for you.

Appendix Q: English Language Test Papers in the First Semester of 2015/16

CRC English First Semi.for final Exam. for Grade 7th in 2008 E.C.

Name -----No-----Section-----

- I. If the statement is ``true``say ``true``or if it is false``say ``false``
1. Verb is show an action.
 2. The past tense of buy is bought.
 3. We play foot ball in the house.
 4. The plural form of ``sheep``is sheep.
 5. Hospital is a place where teachers work.
 6. Un countable nouns have no plural forms.
 7. A person who catches fish is called doctor.
 8. The word expensive and chear have the same meaning.
 9. Comparative adjective is comparing two things.
 10. The comparative form of heavy is heaviest.
- II. Match the word column ``A`` to answers column ``B``
- | ``A`` | ``B`` |
|--------------------|------------------------------|
| 11. Comparative | A.Comparins two things |
| 12. Family members | B.Father,mother,son,daughter |
| 13. Urban | C.is playing |
| 14. Rural | D.in country side |
| 15. Present tense | E.in the city |
- III. Choose the best answers for the following questions.
16. Sugar is-----than salt. A.Expensive B.most expensive C.more expensive D.all
 17. June,July and August are-----seasons. A.Spring B.Autumn C.Summer D.Winter
 18. ---is a word that names of person,power and thing. A.Verb B.Adjective C.noun D.none
 19. He goes school-----foot. A.by B.in C.on D.at
 20. One of the following word is ancautable ?. A.less B.more C.much D.all
 21. A person who sell meat is-----, A.merchant B.shop keeper C.butchers D.tailors
 22. If you get birr from bank,medicine get from . A.shop B.church C.Hotel D.Hospital
 23. The opposites of word village is----. A.country side B.House C.Town D.none
 24. Would you send me-----book for to day. A.your B.yours C. MY D.all
 25. How-----people live in Addis Ababa. A.much B.many C.are D.age
 26. Have you-----your home work yet?. A.do B.done C.doing D.did
 27. My brother had a test-----Sunday. A.in B.for C.with D.on
 28. Gabi is a traditional clothes. A.True B.false C.all
 29. The red fox was----- last Sunday. A.kills B.kill C.killed D.all
 30. Where----- you live?. A.in B.he C.do D.all

Appendix R: English Language Test Papers in the First Semester 2015/16

Arbooyyee CRC 2008 E.C first semester Examination for grade 7

I. Write for correct statement false for wrong statement

1. A comma is used to separate list of items
2. Adverbs are words that describe verbs
3. Letter can be classified in to two catagoies
4. A past lard is a small card used for sening messages through post office
5. Capitalizion is not punctuation

II. Match the word in "A" with "B" their relatives

- | A | B |
|---------------------------------|-----------------------|
| 6. Sweet | A. Slowly, fluently |
| 7. She visits lalibela everyday | B. Delicious |
| 8. He does speak afaan oromoo | C. Neither do I |
| 9. Pronoun | D. So am I |
| 10. Adjective | E. He, she, we, they |
| 11. Adverb | F. Small, long, short |
| 12. Why question | G. Who, where, when |

III. Choose the correct answer from the give altarnatives

13. A: how do you do Bakale B: _____ A. I am fine thank you
B. I am 12 years old C. I am happy to meet you D. How do you do Cirma
14. A what time is it B: _____
A. Yesterday B. it is two o'clock C. tomorrow D. all
15. _____ they sell the horse A was B were C. do D. home
16. Ali is going to lalibela A. Neither am I B. So am I C. So do I D. neither am I
17. You should walk _____ when crossing the roa A. Quickly B. carelessly C. carefully D. A & B
18. He speaks English _____ he cannot write it A. Can B. and C. both D. but
19. Which one is informal letter? A. Professional B. Close friend's letters C. relative letter D. B & C
20. She speaks both Somali _____ amharic A. And B. but C. or D. because
21. _____ is a large mass of land? A. A continent B. A populate C. a land locked D. A & B
22. _____ he play foot ball? A. does B. do C. were D. A & B
23. Ethiopia is _____ in the horn of Africa? A. Situated B. located C. comparied D. A & B
24. Which one is possessive pronous? A. My. Your B. her. his C. our. your D. all are correct

IV. Fill the black space

25. _____ is a word that describe a noun
26. _____ is a word that describe a verb
27. She we they are _____

V. Write the past form of the following words

28. Teach
29. Think
30. Hear

Appendix S: English Language Test Papers in the First Semester 2015/16

AT CRC FIRST SEM. ENIGILSH FINAL EXAMINATION FOR GRADE 7TH IN 2008 YEAR

NAME-----GRADE-----NO-----

I. WRITE "TRUE" IF THE STATEMENTS ARE CORRECT OR WRITE "FALSE" IF ITS INCORRECT.

- 1. Comparatives used for comparing two nouns.
- 2. Indirect speech refers to the actual words said by some one.
- 3. Synonym is a word haring the same meaning.
- 4. Computers do not have brains.
- 5. Conjunction is join two or more sentences.

II. Mach with Their Opposite.

- "A"
- 6. Ancient
 - 7. Attack
 - 8. Expensive
 - 9. Deep
 - 10. Beautiful
 - 11. Borrow
 - 12. Kind
 - 13. Polite

- "B"
- A. Defend
 - B. Modern
 - C. Shallow
 - D. Ugly
 - E. Cheap
 - F. Rude
 - G. Lend
 - H. Cruel

III. Choose The Best Answer For The Following Questions.

- 14. If you are seriously sick you-----go to Hospital. A. would better B. had better C. had D. better
- 15. The little animal hide ----- A. him self B. them self C. her self D. it self
- 16. The Earth ----- around its axis. A. rotate B. rotates C. rotated D. is rotating
- 17. Our Geography Teacher is -----European. A. an B. a C. the D. that
- 18. This is the place -----I was living. A. when B. where C. who D. that
- 19. After I have finished washing dishes, I-----my home work. A. do B. did C. will do D. th
- 20. I go to school every day ----- Saturday and Sunday. A. except B. for C. at D. such
- 21. Take your pen -----write down what I say. A. but B. or C. and D. on
- 22. The School will remain closed ----- tomorrow. A. across B. from C. on D. at
- 23. I am going to -----Addis-----Harar. and/it B. neither/ nor C. either/ or D. none
- 24. Abdi-----won the compitations. A. should B. need C. could D. ought to
- 25. Which Countries share borders with Ethiopia? A. Libia B. Ertrea C. Egypt D. Mali
- 26. The Hotel is not so-----as the theatre. A. large B. long C. quite D. hot
- 27. I-----see you tomorrow at five O'clock. A. will not B. will C. I think D. all
- 28. He is rich-----he is not happy. A. although B. can C. but D. when
- 29. A person who makes suits and coats for men and women say to be-----
A. Post man B. nurse C. tailor D. farmer
- 30. That is your book and this one is ----- A. I B. me C. my D. mine

Perp. At CRC

Appendix T: English Language Test Papers in the First Semester 2015/16

Name ----- Grade ----- section -----

At CRC 1st semester English final examination for grade 7th in 2008 E.c

Say true or false

1. Verbs are words that shows action.
2. A pst card can send message.
3. Debate is forms alimented.
4. Advreb of manner answers the questions when.
5. The use of but to join entrusting ideas.

match the words under column <A> with their meanings under column

A	B
6. one day	A. old
7. job	B. 24 hors
8. batter	C. now
9. not young	D. I'm
10. not two	E. work
11. at this time	F. war field
12.short form of I am	G. one

choos the best answee from the given attentive.

- 13 She speaks ----- English A. and B, both C.or D. but.
14. Holidays are celebrated -----
A. fenr B. yearly C. nears D. day
15. Harer is famous ----- the hyena men.
A. of B. for C. to D, in
- 16.Buru kans Grad morning ? A. Good after noon B. Good night C, good mooring.
Abdata : ----- D. Good bye .3
17. Hello my name is Dureti what is yours? ----- pozia.
A. I am B. she is C. He is D. you are
18. ----- owns the shop? tedese A. whose B. who C. where D. why
19. Bilisume ----- English book every day.
A. carry B. carrying C. carries D. none
20. Ali is going to jimma -----
A. so do I B. so am I C. Neither is I D. none
21. I want to know more ----- Ethiopian culture.
A. so B. non C. about B> all

I. use the words in bract correctible

22. He ----- his homework last period (do)
23. total ----- to America every bay (go)
24. we get our wages ----- (month)
25. coffee is ----- in Ethiopia (grow)

Set by language departmen