Evaluation of the Content Validity of the English Language Test in Erbil Governorate Schools

Madih Asaad Ahmed

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Arts
in
English Language Teaching

Eastern Mediterranean University
January 2018
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

—————————————————————
Assoc. Prof. Dr. Ali Hakan Ulusoy
Acting Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Arts in English Language Teaching.

—————————————————————
Assoc. Prof. Dr. Javanshir Shibliyev
Chair, Department of Foreign
Language Education

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Arts in English Language Teaching.

—————————————————————
Assoc. Prof. Dr. Javanshir Shibliyev
Supervisor

Examining Committee
————————————————————————————————

1. Assoc. Prof. Dr. Emre Debreli   ————————————————————

2. Assoc. Prof. Dr. Javanshir Shibliyev   ————————————————————

3. Asst. Prof. Dr. İlkay Gilanlıoğlu   ————————————————————

ABSTRACT

This study aims to evaluate English language tests of seventh grade at basic schools in Erbil governorate-general directorate in terms of their content validity. More specifically, it tries to find out to what extent the content validity of the tests relate to English language content areas in terms of grammar, vocabulary and pronunciation, and four language skills such as listening, reading, writing and speaking. Hughes (1989) pointed out among the validity types the most relevant one is content validity since it is a means to check the attainment of objects of each content the syllabus of a certain domain.

For the research design, a qualitative method is used and the data was collected through 160 summative test samples for the purpose of this study from the schools. The findings of the study revealed that grammar covers the most part of the tests; whereas both vocabulary and pronunciation were less focused in a manner that they take the second and third places, respectively.

Concerning the language skills, it was found that the test developers focused mostly on writing, and reading stayed at the second stage. However, both speaking and listening skills were totally ignored in the tests.

Overall, it can be concluded that both grammar and vocabulary, as two basic language areas, were mostly reflected the content validity of the English language tests used in the schools of Erbil; whereas, pronunciation was rarely reflected the content validity. Moreover, although both reading and writing skills were adequately reflected the

content validity of the tests, other language skills like speaking and listening were fully ignored.

Keywords: Language Areas, Language Skills, Evaluation, Content Validity, English Language Test.

ÖZ

Bu çalışmanın amacı, Erbil genel idaresi bünyesinde bulunan ilköğretim yedinci sınıf İngilizce dil testlerini kapsam geçerliliği yönünden değerlendirmektir. Daha açık bir ifadeyle; dilbilgisi, kelime bilgisi ile sesletim ve konuşma, dinleme, okuma ve yazma gibi dört temel dil becerisi açısından bu testlerin ne oranda İngilizce dili ile ilişkili olduğunu ortaya koymayı amaçlamaktadır. Hughes (1989), belirli bir alandaki her bir ders programı bileşeninin amaca ulaşmasında bir araç olduğundan kapsam geçerliliğinin, diğer geçerlilikler arasında en ilişkili olduğunu belirtmiştir.

Araştırma yöntemi olarak niteliksel yöntem kullanılmış olup, bu çalışmada kullanılmak üzere okullardan toplanmış olan veriler 160 dönem sonu sınavından elde edilmiştir. Çalışmadan elde edilen bulgular, dilbilgisinin testlerde en yaygın olarak kullanıldığını, kelime bilgisi kısmına ve sesletime en az oranda yer verildiğini, böylece anılan sıraya göre ikinci ve üçüncü sırada yer aldıklarını ortaya koymuştur.

Dil becerileri göz önünde bulundurulduğunda değerlendiricilerin, en çok ikinci aşamada bulunan yazma ve okuma becerileri üzerinde yoğunlaştığı görülmüştür. Bununla birlikte, konuşma ile dinleme becerileri testlerde yer almamıştır.

Genel olarak, iki temel dil alanı olarak en çok dilbilgisi ve kelime bilgisinin, Erbil okullarında kullanılan İngilizce dil testlerinin kapsam geçerliliğini yansıttığı görülürken, sesletimin kapsam geçerliliğini en az yansıttığı sonucuna varılmıştır. Ayrıca okuma ve yazma becerilerinin testlerde kapsam geçerliliğini yeterince yansıttığı görülse de, diğer beceriler olan konuşma ve dinleme becerilerinin tamamen ihmal edildiği görülmüştür.

Anahtar Kelimeler: Dil Alanları, Dil Becerileri, Değerlendirme, Kapsam Geçerliliği,

İngilizce Dil Testi.

DEDICATION

I dedicate this study to my family; first, I would like to dedicate my dear father and mother. Additionally, to my wife, Arkhawan, who worked hard and supported me throughout my studies. Moreover, she faced many problems during my absence to help me achieve my goal. Furthermore, to my sister, Nazanin, who helped take care of my family and saw lots of difficulties during two years of my study. Additional dedications are to my beloved children Diya and Roya, who lived apart from me during my studies. Finally, to my brothers and sisters and all the relatives who supported and encouraged me to complete my studies and achieve this precious certificate.

ACKNOWLEDGMENT

I would like to express my special thanks to my dear supervisor Assoc. Prof. Dr. Javanshir Shibliyev who supported me in order to carry on my study in that difficult times and giving me his valuable feedback to finish my research.

Also, I express my gratitude to the examining committee members Assoc. Prof Dr. Emre Debreli and Asst. Prof. Dr. Ilkay Gilanlıoğlu for their creative feedback to my research.

In addition, I would like to send my deepest gratitude to Assist Lecturer Hawraz Qader Hama in English department at Raparin University for his continuous support and giving me his feedback to complete my study. Moreover, to my nearest and dearest brother and sister Rabar Azad and Tazhan Kamal for their uninterrupted support from the starting point till the end of my study.

Furthermore, my special thanks to the Erbil general educational directorate and its branches for giving me permission in order to collect my data and gathering the test samples.

In conclusion, my deepest appreciation to my beloved wife and two dear children for their patience and continuous support to achieve my goal. Also, I owe my parents, brothers and sisters particularly my elder sister for their sustain help and taking care of my children and my wife. Without their support, I couldn't achieve my MA.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

INTRODUCTION

1.1 Background of the Study

One of the significant processes of teaching and learning program is testing because test is an important way to evaluate both learning and teaching program; and this process has been carried out for a long time. Thus, in order to develop a test, it is necessary to concentrate on the aims of learning and teaching program; also, the students should take the tests in the middle or at the end of their study program or after finishing learning and teaching process (Liesman and Kartio 1991).

In relation to the education system in Kurdistan region government (KRG)-Iraq, in recent years, the Ministry of Education in KRG-Iraq has started to reform the exam system for all the students of basic (primary and secondary) and preparatory schools while taking their tests. The test designers at the Ministry of Education have tried to design a new shape of the tests including both formative and summative tests in both basic and preparatory schools.

The very first step was separating the national exams, which contained 9th and 12th grades from the 4th, 5th, 6th, 7th, 8th, 10th and 11th grades. But for the first, second and third grades students, they do not have summative exams, they have only a few kinds of formative exams like daily quiz and dictation. Regarding the national exams, the students are taking their exams on the same day and at the same hour and then only

one question is used for the students in the whole region and the tests are designed by the test experts in the Ministry of Education. Also, all the questions are multiple choice, which means students are required to answer the whole questions by choosing the right option. Although multiple choice tests seem to be simple, they are very difficult to design.

According to Brown & Abeywickrama (2010, p. 67) "multiple choices are very simple in appearance while they are very difficult to design correctly". However, in KRG, most teachers have not been trained about how to design such questions; and when they analyze their multiple choice tests items, there are many weak points in the design process of the items.

On the other hand, the remaining stages like 6th, 7th, 8th, 9th, 10th, and 11th exams are different from the 9th and 12th grades because the schoolteachers design their tests. However, the Ministry of Education guides schoolteachers on how to design their English language tests and divides the scores as it discusses by showing a table in the next paragraphs.

Since 2003, the KRG has decided to design a new system for education and separated the education system from the Iraqi government. The Kurdistan region project was implemented in 2007, when students, in general, faced many problems particularly in English. The new English language curriculum was called (Sunrise) which is used for both basic and preparatory schools.

Amin (2017) mentioned that before 2007, the English language curriculum in Iraq emphasized grammar, dialogue, sounds and exercises that's why the KRG decided to

develop the English curriculum which reflects the English language areas and the four language skills at that time students become familiar to the language content and break the barriers that they have faced before especially speaking skills.

The new process has changed the whole basic and preparatory schools' curriculums that is why the ministry of education in KRG decided to change the testing design in basic and preparatory schools as well. Educational directorates send guidance every year to the teachers about how to design their tests and divide the scores on each question during the exam in order to help them design the tests more academically. For example, in the decree of the KRG Ministry of Education number (6 private) from 14/12/2016 (see appendix A) it is stated that the English testing scores for the first and second semester of the seventh grades should be divided based on some criteria in a way that 50% of the questions should be related to multiple choice and the remained 50% should be concerned with the writing skills.

In order to have more knowledge, the following table shows how to provide the seventh-grade test scores in a complete year of study which means the first and second semester of 2016-2017.

Table 1: Providing Seventh Grade Scores

First term		Mid exam	
Oral and daily activity	Written	Oral and daily activity	Written
	Multiple choices 50%		Multiple choices 50%
Reading = 3	Grammar = 4	Reading = 3	Grammar = 8
General questions = 4	Vocabulary and pronunciation = 4	General questions = 4	Questions about reading passages = 4
Activities = 3	Functional language = 3	Activities = 3	Vocabulary = 6
			Pronunciation = 6
	Essay questions 50%		Functional language = 6
	Vocabulary = 2		Essay questions 50%
	Spelling = 2		Spelling = 6
	Questions about reading passages = 5		Unseen passage = 10
			Composition = 14
Total = 10	Total = 20	Total = 10	Total = 60

Hama (2015) stated most of the experts believed that multiple choices reduce the trustworthiness of the tests scores, and teaching process may give an ambiguous picture which loses its own value in this way.

As it is known, testing is a significant and necessary part in the realm of teaching and learning process. By taking exams, students can pass from one stage to another; so testing is an important way to decide whether a student deserves to pass or fail from one stage into another. The principal goal of the test is to assess or evaluate the

4

students' success in attaining the essential objective of education, this kind of exam is regarded as a part of content validity. This means that every test should have a high content validity and reliability; also, it necessarily should be practical.

1.2 Problem Statement

This study tries to evaluate the content validity of the final English language exam of the seventh grade in the academic year of 2016-2017. The test items, concerned with the content validity will be analyzed in terms of their content areas and language skills for the students at Erbil basic schools to know if the test actually evaluates what it has to measure in KRG-Iraq. Regarding the context, focusing on such principle as validity, especially content validity is very crucial. Likewise, it is significant to shed light on measurement. For example, Fulcher (2013) stated that measure in language testing refers to the strong 'trait theory' of validity: what we think our test measures is a real, stable and part of the test taker.

Based on the explanations, test evaluation or assessment can be seen as an academic and systematic procedure to assess particular format of the test and to make necessary changes concerning on the outcomes of the evaluation process for the improvement of the syllabus. It is very significant to evaluate the teacher education abilities systematically in the area of testing and to identify their strengths and weaknesses in order to direct them on how to prepare an academic test.

It seems that many practitioners who prepare tests are not equipped with knowledge of how to 'test' their tests on principles. This is especially evident when content related validity is in question. Therefore, it seems to be of utmost importance to focus on this issue.

1.3 Purpose of the Study

This study attempts to evaluate tests of seventh grade at basic schools in Erbil governorate-general directorate in terms of their content validity. It also tries to find out to what extent the English language tests of seventh grade at Erbil governorate public and private schools refer to their language areas in terms of grammar, vocabulary and pronunciation, and then to find out whether the tests have any relation to the topics that have been taught in terms of four language skills such as listening, reading writing and speaking.

1.4 Research Questions

To this purpose, the study attempts to answer the following research questions:

1. **To what extent do the English language tests used at Erbil basic schools reflect the content of the materials been taught in terms of the following language areas?**

   a. Grammar

   b. Vocabulary

   c. Pronunciation

2. **To what extent do the English language tests used at Erbil basic schools reflect the content of the materials been taught in terms of the following language skills?**

   a. Reading

   b. Writing

   c. Speaking

   d. Listening

1.5 The Significance of the Study

The reason behind choosing this topic is that there seems to be no other studies about analyzing and evaluating tests through content validity before at Erbil governorate basic schools. The findings of this study are expected to provide useful feedback to the English language supervisors and test designers about the effectiveness of tests that are used by teachers. Moreover, the findings will lead to make possible changes in test format that are given by general educational directorate in Erbil/Kurdistan Region-Iraq.

1.6 Definition of Terms

Test: "refers to evaluation of person's ability, knowledge or performance in a given domain" (Brown 2004, p. 3). It means that the process of testing can evaluate the students or test-takers capacity and also, this process will decide whether the students deserve to pass from one level into the highest one or to fail which means to stay at the same level without proceeding.

Summative test: Students will take this test at the end of the semester or year so as to evaluate the ability of a student on what he/she has learned during class (Hughes 1989).

Validity: Brown (2004, p. 29) defined validity as "the degree to which a test measures what it claims, or purports, to be measuring". Which means if the test has the capacity to measure or the test totally measured that it is valid.

Content validity: Fulcher and Davidson (2007, p. 6) defined content validity as "any attempt to show that the content of the test is a representative sample of the domain to

be tested". It means that if the test is related to the topics that have been taught, the test has content validity; so the questions should be related to the subjects.

1.7 Summary

This chapter discussed the background of the study, problem statement, purpose of the study, research questions, significance of the study, definition of terms, and summary of the study. Chapter two will discuss the review of related literature which contains the theories and related studies.

Chapter 2

LITERATURE REVIEW

Some theories and related studies are demonstrated in this chapter which underlies the research. This chapter aims to discuss the test and the five principles of testing. In addition, assessment and its components will be discussed. Moreover, different test samples of seventh grade will be shown in order to analyze the items according to the requirements of the study.

2.1 What is a Test?

Many experts have defined test according to their understanding, most of the definitions have a similar shape but the expressions are different. For instance, Heaton (1988) said that a test is made due to support learning and encourage the students or predominantly it is related to the evaluation of the test takers' performance. Moreover, it presents an aid in order to define the character of the examinees. Besides, Bachman (1990) stated that a test is an assessing tool produced with clear procedures to show person's performance through a particular example.

In addition, Institute for Curriculum Development and Research (ICDR) (1999) pointed out that a test is an organized process of defining the level to which instructional goals are achieved by students. Finally, Good (1973) defined a test as the way to evaluate the capacity of students by giving them different types of questions due to answer, the aim of this process is to measure the students' ability.

Based on the definitions that have been mentioned by the scholars above, a test can be introduced as a systematic process or method or a tool which provide a group of questions and the questions asked through oral or written that are basically formed to evaluate student's capability.

2.2 Formative and Summative Tests

The formative test is conducted when the teacher or the lecturer is teaching in the classroom. In this process, the development of the students is monitored during teaching process in learning certain materials. The formative test highlights on evaluating all the anticipated results of the unit instruction. This test helps the teachers to get a beneficial outcome in order to develop the learning methods through reviewing the instruction and creating the materials more effectively, for instance, dictation, quizzes and midterm test that is done by the instructors inside the classroom (Bandoro, 2014).

On the other hand, Brown (2004) stated that summative test is another kind of a test which is related to the degree to which the test takers have gained at the end of semester or learning process. This test summarizes the whole study that a student has learned during one semester or an academic year of study, which means everything will be included in the exam from the beginning till at the end of the course. Students may prepare themselves for this test in order to get a high mark and stepping up to a higher level. For example, final exams and general proficiency tests are accepted as a sample of the summative test. In addition, most of the points of a test in one semester emphasize the final test; it means that getting a good point in this test makes the way easier to the test taker to succeed from their current level into the highest one. Finally, the test decides if a student deserves to pass or fail.

2.3 Types of Tests

Different types of tests are required via various intentions. Thus, based on the purpose of testing, tests are categorized into different kinds. Brown (2004) classified tests into five types as achievement tests, diagnostic tests, placement tests, proficiency tests and aptitude tests.

The first one is achievement test. This kind of test is referred to learning and teaching; it means in general, achievement tests are correlated to the syllabus- content of the test. It is evaluating the skill of students in the specific syllabus. Also, it is often summative since it is taken at the end of the end of course or term of study. A specific approach to teaching and learning should be reflected by a good achievement test that is been adopted (Heaton, 1988).

Hughes (1989) believed that achievement test is definitely fixed to syllabus contents and classroom understandings. Likewise, Weir (1993) mentioned that meanwhile achievement test can be seen as an essential part of learning process, it works best if it contains items and tasks about which students are familiar to. Such test takes place at the end of semester or school year for defined purposes.

The purpose of taking achievement test is to assess student's development in language skills and to know to what extent they are in relation to the syllabus that they have been studied before (Harmer, 2001). Also, Weir (1993) believed that the fundamental goal of an achievement test is to know to what extent the students could understand the language that was studied during the course and in the way they used this language receptively and productively. In brief, achievement test is taken to know the percentages of students are successful in learning the language and how much the aim

of the language course is attained. From the achievement tests, student's scores can provide the related information to teachers, parents, curriculum developers, students themselves and so on (Bandoro, 2014).

The second one is diagnostic test. Diagnostic test is related to the learning difficulties, it helps both lecturers and students to be familiar with those problems that they have faced with the language (Thorndike, 1977). This test seeks to find out the basic reasons of the learning problem. For instance, what are the basic reasons making students confuse present tense and present perfect tense? The goal is to find out the real causes of learning problems and to develop a strategy or plan for remedial action.

The third one is placement test. This test is conducted at the beginning of the study before starting the real course in order to measure the test takers performances. Harrison (1983) defined placement test that is designed for the new students when they want to start a new course, and this test evaluates their skills through getting the grades then they can start their course by measuring the placement tests outcomes. It is not such a kind of syllabus but it is always evaluating the candidates' capability of their language learning rather than a particular point of learning. Finally, placement test supports to get an obvious image of test takers activity on a course before starting the program.

The fourth one is proficiency test. The proficiency test is designed to evaluate the language skills of the test takers capability without any instruction that they have taken before which means it is not referring to any program of teaching (Hughes, 1989). The aim of this test is to specify the candidates' ability to use language in a test in order to use the outcome of their upcoming activity. Currently, proficiency tests become an

obligatory test because most of the well-known universities around the world will ask such proficiency tests as IELTS and TOEFL of those people who apply to start studying at those universities.

The fifth one is aptitude test. This kind of test is designed to evaluate an overall ability of the candidates first and then, they can take their course. This test is designed in order to learn any language in the classroom. Aptitude tests are divided into two standardized tests; the first one is (Modern Language Aptitude Test MLAT) and the second one is (Pimsleur Language Aptitude Battery PLAB). These two are English language tests and students are required to take these tests in order to perform their language tasks; for example making differences between speech sounds, showing grammatical functions and so on. Both tests are taken in the United States (Brown, 2004).

2.4 Test Item Types

Gronlund (1982) divided test items into two main types; the first one is supply-type and the second one is the selection-type. Both types are described as follows:

1- Supply-type items (Test taker supplies answer)

The supply-type is such test items when the test takers write their words due to respond the question which has been written in the test. At that time, the atmosphere is not easy to control the test takers' response. The supply-type items are divided into four:

A. Essay – extended response

The extended response essay is related to the general question; it gives permission to the test takers to respond their questions without any limitation which means they can answer their questions as wide as they can.

Test developer designed such a story.

For example, in this question, the teacher asks the test takers to answer the question between the brackets in a paragraph. The questions between the brackets have involved the name of a person, setting, time and mission.

If the question relates to a person, the test taker needs to think about the character he/she would like to include in the story. The name of the character should be recognized and the person needs to be described how he/she acts and looks.

Who is in the story? _____

At the starting point, the setting of the story needs to be identified. It may be a city, village, house, historical place, coastline and a time frame might be involved also. At that time setting needs to be described, for example, what it looks like and any attitude needs to be suggested.

Where does it take place? _____

Draw the major events of the story briefly not in detail.

What happens? _____

The examinee should think how to finalize the story.

How does the story end? _____

Hence, the examinees should write according to the information above.

(Source: seventh grade sample 2016-2017). See Appendix (B)

The guidelines above advise the test takers to write an essay in paragraph according to their opinion by creating an image of the topic in their own mind. The item guides the students how to write paragraphs during essay writing while the content of the paragraphs refers to the views of the test takers in their writing. The scoring of this kind of test is focused on the questions that have been asked by the examiner in order to describe the benchmarks in which the writing types need to be evaluated.

    B. Essay – restricted response

In the restricted response, the test takers cannot write as much as possible; thus, their writing has been limited in length, generality and the arrangement of the examinees' response. For instance:

Instruction: Based on the passage above write an email to your pen friend about what you like or don't like doing in your free time. (No more than one paragraph).

(Source: seventh grade sample 2016-2017). See appendix (C).

C. Short answer or word phrase

This kind of response will be answered by the test takers in brief. It is also limited completely to the number of known results. Based on the following sample the examinee reads a short passage about the village hotel and will answer the question in sequence according to his/her understanding the passage in brief. The examiner asks:

Instruction: Read the passage and answer the questions in brief.

1. What is the place name for young people?

2. How many restaurants are there?

3. Where can you see films?

4. How many rooms does the hotel have?

5. Where can you dance?

(Source: seventh-grade sample 2016-2017). See appendix (D).

D. Completion (Fill in the blanks)

From this type of test, it is required to the test taker to fill the spaces in the correct one or choose the correct words that are missing such as:

Instruction: Choose the correct word to fill in the gaps.

1- Anna can _____ the guitar (plays, play, playing).

2- _____ does the film start? It starts at 10 o'clock. (where, when, what).

3- Banaz _____ 20 kilometers last weekend. (walk, walked, walks).

4- Pedro doesn't _____ banana. (like, likes, liking).

5- What do you like doing _____ Saturday (in, on, at)

(Source: seventh grade sample 2016-2017). See Appendix (E).

In this type of test, the examinees should choose the correct response from both categories either filling the blanks or choosing the correct answers. The scoring totally looks like short answers item type.

2- Selection-type items (Student selects answer)

There is unlimited structuring provided by selection-type items in order to evaluate different learning consequences through simple to complex, meanwhile the examinees have no right to define the problem once again, deliver partly correct responses, or show learning unrelated to those required via the test items (Gronlund, 1982).

This kind of test is divided into three test item types described in the following examples:

A. True-false

The test takers in this kind of test should demonstrate whether the given test is true or false by reading a short paragraph such as:

Instruction: Read the sentences and mark them [T] for true and [F] for false.

1. Banaz is a teacher.

2. She usually gets up at 7.

3. She is 12 years old.

4. She doesn't wash her teeth.

5. She goes to school by bicycle**.**

(Source: seventh-grade sample 2016-2017). See appendix (F).

Through the instruction of true-false item type, the test takers recognize the sentences that have been given to them from the exam if a statement is true they have to write (T) or false they have to write (F).

B. Matching

In matching exercises, two types of sentences will be presented in opposition. The second side is the completion of the first side hence the examinee should match the first part sentences into the second part in order to create them in one sentence, for example:

Instruction: Match the words.

A	B
Space	bag
Washing	pool
Sleeping	station
Sun	machine
Swimming	glasses

(Source: seventh-grade sample 2016-2017). See Appendix (G).

C. Multiple-choice

This kind of test brings such a headache to the examinees and it provides lots of answers. At that time the examiner gives around three or four options to the test takers and they have to choose the correct one.

Instruction: Choose the correct answer from the following options.

1. She is ………………. her toys. (A- touch B- touched C- touching)

2. He ………………. got any books. (has, hasn't, haven't)

3. Tom ………………. run very fast. (can, could, should)

Another version of multiple choice such as:

Instruction: Choose the fruit one from the following options.

A) Avocados

B) Carrots

C) Celery

D) Radishes

E) Green onion

(Source: seventh-grade sample 2015-2016). See Appendix (H).

In multiple choice item, the examiner needs the examinee to choose the right answer in order to complete the sentence. The response options are given through the textbook. The examinees must choose the correct one. The scoring is objective in this item test because the test takers have only one option and the item test has only one true response.

It is worth to say multiple choice test items (MCTI) have advantages and disadvantages, at the beginning some pros will be discussed as it is showing in the

following samples. First of all, MCTIs are more flexible than other item types for evaluating a variability of content and instructional objects. According to Mousavi (2009), MCTIs are chosen by assessment specialists because the items' sampling of content is usually superior when compared to other item types.

The second one, it is easy to administrate the MCTIs which means that a large number of options can be given to the examinees in an individual testing session. This supports the examiners to include a great number of various tasks or single objects in the testing session (Harris 1969).

The third one, MCTIs do not permit the test takers to use the strategy of avoidance which means to avoid the correct response or difficult problems via prolonging their answers such as in composition writing. The examiners in multiple-choice exams will not face such a problem in the area of scoring which they will not be affected by the personal judgment. Burton et al. (1991) believed that if MCTIs are well designed at that time they don't have such a power to guess and subsequently fit for creating more reliable scores.

The last one, scoring in MCTIs are easier than other items, they don't take lots of time, thus, help the examiner to score the tests in a short time which means saving time and the scoring can be done in variable ways by using different kinds of instruments, for example, computers and machines without using hands (Clegg & Cashin, 1986).

On the other hand, the cons will be discussed based on the next samples:

It is clear that currently multiple choice test items are used broadly by testers in the exam papers. Conversely, their usefulness is limited. It looks like other test items teachers should be aware of its drawbacks (Hama 2015). Firstly, in order to design fruitful MCTIs, it takes lots of time. The reason behind this is that the test developers will face difficulties in finding whether the test items are distractors. According to Mousavi (2009, p. 432) the nature of MCTIs does not depend on what is tested, but the ability of item writers to construct well-functioning distractors.

Secondly, estimating the exact response possibly will have a significant influence on the scores of the test. Thirdly, MCIs are utilized to test merely recognition knowledge (Hughes 2003). In multiple choice items the test takers emphasis only on the options more than creating an answer. These items are not appropriate to evaluate some other learning results like expanding, summarizing, giving samples, and saying individual opinions. Fourthly, Backwash; according to Mousavi (2009, p. 47) "in order to find more information on Backwash" possibly has harmful influence. This means the examinees in such a kind test will focus on guessing not on their language capability. Furthermore, they concentrate on the memorization without focusing on the language learning. Also, it is a fascinating way for cheating. In this way, the students can communicate by using their body language (non-verbally). Lastly, the multiple choice test items usually are not acceptable by the language teachers because they rarely reflect the real life (Mousavi 2009).

2.5 Continuous Assessment

Assessment systems are divided into two major parts by formal educational settings. The first one is continuous assessment and the second one is called end term assessment systems. Continuous assessment is characterized by researchers from an

instructive point of view in the accompanying ways: Continuous assessment is a student assessment framework that works at the classroom level and is incorporated with the instructional process (ICDR, 1999). Likewise, Yoloye (1984) defined continuous assessment as a technique for assessing the advance and accomplishment of understudies in instructive establishments.

Thus, from these perspectives, we can comprehend that continuous assessment is progressing kind of evaluation which encourages learning as it frequently involves students in work by utilizing different procedures of assessment. Facts demonstrate that today schools and colleges are underlining the utilization of continuous assessment for it has different favorable circumstances in teaching learning process (ICDR, 1999).

In general, continuous assessment system brings tutors and learners together in a helpful attempt to attain the aims of classroom instruction. It checks student's development and the efficiency of instructor's instructional methodologies (Asmare, 2008).

Researchers like Weir (1990) said that for better development of tests, it is unequivocally prudent to determine the goals of the course ahead of time. Conversely, all sought goals of a course can't be measured with paper-and-pencil tests like concept acquisitions, psychometric aptitudes, oral communication, and so forth. Such content areas of a subject can be assessed by using continuous assessment system, because continues assessment system includes assessment techniques like observation. For instance, the best way to evaluate speaking skills is observation which includes well-set checklists.

ICDR (1999) noted that it is the schools' responsibility to decide the weight of continuous assessment and end term assessment. It is recommended that the percentage of this test should be 50. To illustrate, weight needs to be collected through various techniques of continuous assessment, for example, portfolio, written test, homework, classwork, project, observation, etc.

Finally, ICDR (1999) emphasized that continuous assessment is similar to end term assessment, and it should give a relative opportunity to the inspecting of targets or substance of a given course. So, both tests are accepted to the test contents and they are demonstrated in a syllabus that is dealt with in the classroom.

2.6 The Relationship among Teaching, Testing and Assessment

Teaching and testing in language are seen as a processes that cannot be separated from each other. Both of them are two significant ingredients in the process of teaching and learning. For supporting this notion, Heaton (1988) noted that it is very difficult to work on each component in isolation without mentioning the other. Similarly, Dejene (1994) described testing as an instrument in order to create the best way to the whole process of teaching and learning activity.

In addition, Venkateswaran (1995) pointed out that tests can give lots of information to the teachers and students and that information can have a positive impact on the process of teaching. He insisted that tests support the students as well in order to create an appropriate atmosphere within the class to become proficient in their language and promote learning through their diagnostic atmosphere.

Teachers can achieve much information from tests. For responding the important questions teachers need to focus on good tests; for this reason tests help the teachers

and they ask to what extent their teaching styles have been effective. This question supports the teachers to identify their efforts that have been practiced in teaching.

Teaching and assessment are two correlated things that are completing each other; it appears that, in reality, both characters have a great impact on each other: The assessment methods affect teaching in the classroom activity (Cheng & Wall, 1997). In addition, teaching is the way which evaluates a learner's learning capability, it is not only related to a particular area of study but also adjusting instructional techniques with a different kind of activity or performance in the class (Mousavi, 2009). As follows, each language skill in relation with teacher's observance performances can be accepted as an assessment part of teacher in teaching activity. Brown (2004) discussed the link among testing, assessment and teaching, he emphasized on the distinctions of language practice by showing a diagram. In Brown's diagram four characters of teaching, assessment, measurement, and tests were overlapped due to the assessment of learners' ability. For this reason, currently the instructor gives instruction and assessment to the test takers to evaluate them on how they acquire and yield the topics that they have learned in their classes. Hence, it is possible to say there is a sequential relationship among assessment, measurement and tests with teaching.

Brown (2004) stated that there is misunderstanding about the term testing and assessment; what's more, people believed that these two words are synonyms. But in reality, they are opposite. To get extra information about testing and assessment both terms will be discussed in the following paragraph.

Mousavi (2009) defined assessment as a process of estimating the skill or proficiency of a person. Moreover, Hughes (2003) pointed out that an assessment is the way when

teachers are monitoring their students since they know about the development of their students to see to what extent their students have learned and use a new information in order to change their next plan. In contrast, test is a method for evaluating the capability, performance and knowledge of a person through the result that has been taken by the test takers (Brown, 2004). On the other hand, some researchers believed that there is not any difference between test and assessment; for example, Clapham (2000) stated that no distinction has been seen between both terms.

2.7 Qualities of a Good Test

Tests have a great influence on the quality and quantity of learning. It is also necessary to understand the principles of testing and how they can be applied in practice. It is relatively straightforward to introduce and explain the desirable quality of a good test. Brown (2004) discussed five principles that can be introduced as the qualities or characteristics of a good test namely; practicality, backwash, reliability, authenticity and validity; referring to the favorable effects testing can be beneficial on teaching and learning if it contains the four main principles of quality test.

2.7.1 Practicality

According to Harrison (1983), the practicality in testing relates to the consideration of financial ability, administrative capability and scoring ability of the aim setting. Practicality shows the economy of time, effort, and money in testing. In other words, a test should be easy to design, easy to administer, easy to mark and easy to interpret the results. Weir (1990) believed that practicality, reliability and validity are relying on each other, tests can be accepted valid and reliable if they are practicable. Also, practicality relates to the term of cost, time, administration, and scoring evaluation. For this reason, Brown (2004) defined practicality in four terms cost, time, administration, and scoring or evaluation.

1. Cost: for conducting a test it doesn't need to have an expensive cost, it is very necessary to have a balance between the budget and the cost of the test and does not conduct such a test that needs lots of money.

2. Time: is another significant point in testing which focuses on a proper time to conduct a test. Moreover, the test should not be too long or too short.

3. Administration: avoid of conducting a complex test so the tests should not be complicated, whereas it needs to be quite simple.

Furthermore, Bachman and Palmer (1990) explained that practicality needs three types in order to be assessed. The first one is human resources that contain, for instance, the person who writes a test, scores and the one who administrates a test and technical support as well. The second one is material resources which include room for the test takers, equipment such as computers, prints, etc. The third one is time which includes time development from the beginning of the test until writing the test scores and time for specific tasks such as designing, writing and so on.

2.7.2 Washback

According to Bailey (1986) when the process of learning and teaching has been influenced by testing, it is called as washback. The term 'backwash' is used to refer to the way a test influences teaching materials and classroom management (Hughes 1989). Messick (1989) pointed out the idea of washback is brought from consequential validity. Consequential validity: covers all the consequences of a test, as along with such consideration; its accuracy in measuring the intended criteria, its influence on the examinees during the preparation, the impact on the learner, and the social consequences of a test's interpretation and use (Brown, 2004). Also, he discussed some points in a test that give beneficial washback such as:

- A positive impact on the style of teaching.

- A positive impact on the style of learning.

- Gives a chance to the learners an adequate preparation.

- Provides feedback to the learners in order to progress their language.

- It's introduced to be formative than summative.

Furthermore, Morrow (1986) claimed that the validity of a test relates to a positive washback, while the test doesn't have validity if it has a negative washback.

The notion of washback is related to the things that the examiners and examinees do which it was not required but they do it for the sake of test (Alderson and Wall, 1993).

Washback shows the test is a negative (harmful) or positive (beneficial). Negative washback is said to occur when a test's content or format is based on a narrow definition of language ability and so constrains the teaching learning context. Davies et al. (1999) proposed the following illustration: If for example, the skill of writing is tested only by multiple-choice items then there is great pressure to practice such items rather than to practice the skill of writing itself'. Positive washback is said to result when a testing procedure encourages 'good' teaching practice, for instance, an oral proficiency test is introduced in the expectation that it will promote the teaching of speaking skills.

2.7.3 Reliability

Another important criterion for the quality of a good test is reliability. Brown (2004) defined reliability as a consistency of measurement. Alderson et al. (1995) stated that

reliability refers to the degree to which test scores are reliable when a gathering of understudies takes it in two events at about a similar time.

Thus, reliability is the degree to which a trial, test or any measuring strategy demonstrates a similar outcome on reiterated trials. Instead of the agreement of free viewers ready to reproduce investigate systems, or the capacity to utilize look into devices and procedures that create reliable estimations, analysts would not be able attractively to reach inferences, detail speculations or make asserts about the generalizability of their exploration.

For further discussion about reliability, Fulcher and Davidson (2007) concentrated on three points teachers need to make in classroom assessment, one of which is (test-retest) at that time a similar test is directed twice and a relationship figured between the scores on every administration or it shows the reliability of the test when the test scores come from the similar sorts of the test. The second one is parallel forms in which two types of a similar test are created with the end goal that they test a similar construct and have comparable means and differences. The relationship between the scores of the two frames is taken as a measure of reliability. And the last one is split halves when a particular test is administrated, part of the items is taken to indicate one type of the test and associated with the things in the other portion of the test. The relationship coefficient is taken as a measure of reliability.

There is a strong relationship between reliability and validity. Weir (1993) stated that validity is the first step when someone would like to design a test. On the other hand, to approve that a test is valid, it should be reliable. In the same way, Underhill (1991) believed that reliability is a particular type of validity. A test cannot be accepted as

valid if it is unreliable. In addition, Alderson et al. (1995) explained that if a test is not reliable it cannot be valid; nevertheless valid test seems not to be reliable due to the reality that can be seen outside the test itself. Such realities are testing situations, graders and so on. Moreover, validity works with instruments or methods while reliability works with the results. According to these researchers' perception, it's possible to say a good test should include both validity and reliability.

According to Brown (2004), there are lots of reasons which influence the reliability of a test. These reasons refer to the test itself, test raters, testees and test area conditions. Every reason is demonstrated as follows:

- Test itself: incorporates a length of the test too long or too short, guideline clearness, sufficient examining of test items, etc.

- Testees: relate to the problem that students faced on during the examination time such as illness, motivation towards a test, without sleeping throughout the night and so on.

- Test rater: consists of individual techniques of scoring and so on.

- Test area conditions: contains bright, speaking high, high and lower temperature, test administrators and so on.

Hence, to create the reliability of a test, testers ought to try to decrease the reasons that are mentioned above. Assuming this is the case, test scores can be viewed as an actual reflection of the test takers language proficiency level.

2.7.4 Authenticity

The idea of authenticity in applied linguistics appeared at the end of the 1970s, a growing need had been seen at that time in the area of language teaching and testing (Authentic life). Bachman and Palmer (1996, p. 23) defined authenticity "the degree

of correspondence of the characteristics of a given language test task to feature the features of target language task". They also proposed an idea in order to identify the mentioned target language tasks and altering the test items that have validity. In addition, an authentic text is known to expand the reality of language that has been produced by a real speaker or writer for a real audience and designed to transfer an authentic message of the different category (Morrow, 1977). Moreover, Nunan (1999) pointed out authentic materials such as spoken or written language data is made through a genuine course of communication, and not only written has become the basic aims of language teaching.

Widdowson (1978) debated on the nature of authenticity and offered the distinction between 'genuineness' and 'authenticity', he believed that genuineness can be related to the feature of the passage itself in a real quality, he saw genuineness as a quality of all texts. However, authenticity is defined as a character which shows the relationship between the real life of the study and the reader and which is related to a suitable reply, an attribute by a given spectator which means authenticity as an attribute 'bestowed' on texts by a given audience.

The difference between genuine and authentic language was not completely approved, and the conversation stayed unsolved.

Brown (2004) proposed some features for a test to be authentic:
- The language in the test is as natural as possible which means using a simple language due to being understandable for the test takers.
- Items are contextualized rather than being isolated, items should relate to the context not coming from different contexts.

- Topics are meaningful (relevant, interesting) for the learner which means learners need to be familiar with the test topic, also the tests should have validity.

- Some thematic organizations to items are provided, for instance, through a storyline or episode which refers to the situation.

- Tasks represents, or closely approximates real-world tasks, tests should be taken from the real life of the students and the repetition of their daily life or their class daily life.

In language teaching and assessment, authenticity is usually a negotiable notion that has been discussed in a variable way via the stakeholders. There is no doubt that most of the experts in language testing believed that authenticity is not an easy concept to identify according to their apprehension of language and learning. Additionally, Lewkowicz (1997) said the experience revealed that the concept of authenticity is not so easy to define in detail.

2.7.5 Validity

Many experts have defined validity according to their points of view. For instance, Weir (1993) described validity as the degree to which a given test or any of its component parts ought to measure what is claimed to measure, not anything else. Likewise, Hughes (1989, p. 22) defined validity as realizing if a test "measures accurately what is intended to measure" which means if the test has the capacity to measure or the whole test measured is valid. Additionally, Thirumalai et al. (2012) believed validity means that the test ought to measure the same for which it has been devised. For example, a test is valid if it tests grammar only which means measuring only grammar use and nothing else. Thus, the test is valid in terms of grammar.

According to Hamavandy and Kiany (2014) validity can be defined as a relationship between the test and the criterion which has been defined before.

According to Davies and Elder (2005, p. 795) "the notion of validity refers to the value of a test and its scores, both of them are strong and precarious". The idea of strong relates to power because all the features of language testing will be controlled by validity. Precarious since it replies the four difficult challenges.

Firstly, the appeal to logic, in the first step it emphasizes on the term of validity and using this term in philosophic logic. Validity can be described as an old son; it is used by the philosophers. Angeles (1981) stated that a logical argument can be seen as valid when its outcome refers to the validity principles or premises; if the argument premises may find it true, then the result is also true.

The second challenge is the claim of reliability. Based on validity, the test needs to have a truth-value or it needs evidence in order to be valid but if the test shows its consistency it has reliability. The link between validity and reliability is not looking like the relationship between form and meaning. Meaning is focusing on the matters, not the forms and it is also disappearing. Similarly, validity is giving life to the test, its uniqueness as a measure, but to exist as an entity, it needs reliability. For this reason, Lado (1961) said that reliability is not specific but it is general.

The third challenge is local and universal. Lots of arguments focused on the humanities and social science in the last period between the enlightenment or universals and the local or relativist opinions. The humanity (and experience) can be understood in similar ways and the relativist (or local) view that contexts are not just apparently but

32

fundamentally different. These arguments relate to a long time which cannot be accepted current. Based on the validity of a test these methods describe the notion of positivist and interpretivist (Lynch, 2003). Positivist validity focused on the relational information between test and an arranged criterion. Their question was "Are we measuring the relevant construct? (Lynch, 2003, p. 151). In contrast, the interpretivist validity said that: What is the test that we are testing? They didn't focus on reliability.

The final challenge is the unitarity and the divisibility. Davies and Elder (2005) pointed out validity gives the extent of the real evidence about persons, regarding the circumstances of using it, currently used as a unitary idea. Due to letting the examiner create a statement about the validity of the test, it shows that it is required to concentrate on some approachable methodology of inquiry. Conversely, validity remains as an issue of belief, a magical type.

It is very important that each item of the test should be tested in an obvious way which means the examiner ought to bring meaningful items in his/her test to the test takers since they don't face such a problem in using their own language during the examination. It is possible to say that test is valid. Validity is divided into five basic types. They are called face validity, construct validity, criterion validity, consequential validity and content validity. All of them will be explained in the following paragraphs.

2.7.5.1 Face Validity

Face validity is related to the ways how the examiners and the examinees think about the test (Harrison, 1983). Moreover, Underhill (1991) noted that face validity is a feature of validity attained through the questions such as on the face of it, does it relate to realistic test? Does the person who takes the test believe that this test is reasonable or good?

There is a similarity between the face and content validity, both of them are determined by a review of the items and not depend on the use of statistical or scientific analyses. For example, you can measure grammar indirectly (Brown, 2004).

Also, Brown (2004) has shown the difference between face validity and content validity. Face validity is not investigated through formal procedures. Instead, anyone who looks over the test, including examinees, may develop an informal opinion as to whether or not the test is measuring what it is supposed to measure.

Although Bachman (1990) stated that face validity is not an actual type of validity, it is a desirable feature for many tests. If a test lacks face validity, examinees may not be motivated to respond the test items in an honest or accurate manner.

Face validity means that students perceive the test to be valid. So, it is highly valid in following conditions (Brown 2004):

- The format needs to be well-organized and have familiarity with the tasks (students feel confident if they faced well).
- Giving an appropriate time for each test (examinees don't feel anxious about the duration of the test).
- Items need to be clear not ambiguous (test taker feel optimistic).
- Clear directions (it is easy to find).
- Tests should be related to their course book (content validity).
- And a difficulty level that presents a reasonable challenge.

Finally, the information above revealed that face validity is not seen as an empirical validity. It is normally working with the appeal of the test to examiners, examinees,

family etc. Thus, the specialists do not need to evaluate or judge for. It works with the surface stages of a test, for example, item numbers, formats and so on (Asmareh, 2008).

2.7.5.2 Construct Validity

Bachman and Palmer (1996) defined construct validity as the ability that proposes the standard for a given test and for comprehending scores that have been derived from the test. Additionally, Brown (2004) stated that construct validity is any theory or method helpful to measure the test, otherwise we cannot get the result of construct validity without using techniques or methods. Through a measurement viewpoint, Kaplan & Saccuzzo (2012, p. 135) defined construct validity as "the agreement between a test score or measure and the quality [or construct] it is believed to measure".

The validity inquiry discusses a framework and it searches in order to create construct validity through concentrating on variable references of evidence. Messick (1989) mentioned that construct validity achieves evidence through content and predictive validity. In another way, Hughes (2003) pointed out content and criterion validity will give an evidence to construct validity.

Fulcher and Davidson (2007) believed that self-esteem and motivation have been made by psychological concepts. For example, English language proficiency is a construct validity especially in speaking test; in it, the interlocutors may focus on fluency, grammar, pronunciation, etc.

Ebel and Frisbie (1991) argued that the term construct relates to a psychological construct, a theoretical notion about a characteristic of human behavior that cannot be measured or perceived directly.

For a term to be a construct, it should have more than one properties: such as being measurable and having a relationship with other different constructs.

Construct validity deals with these notions:

- The measure will perform according to the related theory.

- The scores of the test are explained psychologically.

- Psychologically construct validity motivates the test.

- The test exactly makes a contact of the theoretical construct as it has been explained.

There is no particular way in order to study about measuring construct validity. In most cases, construct validity ought to be shown from various points of view. For instance, imagine an oral interview whose scoring analysis includes several factors such as fluency, pronunciation, grammatical accuracy, vocabulary use and sociolinguistic appropriateness.

Fulcher and Davidson (2007) mentioned that intelligence; love, achievement, attitude, fluency, empathy, and so on refer to the construct validity.

In general, two or more properties need to become construct validity. Firstly, it should be described in a measurable way. Due to measuring fluency it is very significant to make emphasis on the speech of the speaker and so, the interlocutor will decide about the speaker whether he/she is fluent or not. It demonstrates that most of the people will

define fluency in different ways by concentrating on a normal speed of dialogue, lack of hesitancy, because hesitation can be described as a part of construct validity. Secondly, having a relationship with defined constructs and other constructs in a different way. For instance, if we describe both anxiety and fluency, it is clear that anxiety always increases while fluency decreases (Fulcher, 1996).

A test's construct validity involves a systematic collecting of evidence showing that the test actually measures the construct that it was designed to measure.

Lastly, in construct validity, two points are mentioned. The first one is the way how to use the language areas and the four language skills in a given test (Duran, Canale, Penfield, Stansfield, & Liskin-Gasparo, 1985). And the second one is the concept that construct validity mentioned before the test psychologically motivates the test taker (Ebel & Frisbie, 1991).

2.7.5.3 Criterion Validity (external validity)

Criterion validity demonstrates to what extent students' scores are related to other criteria that reflect the same construct. Or it utilized to predict the current or future performance. It connects test outcomes with another norm of interest. It needs independent approach from outside and compares with a test which has criteria due to obtaining evidence (Fulcher & Davidson, 2007).

According to them, criterion validity is divided into two major types, which are concurrent and predictive validity:

- **Concurrent validity**

The concurrent validity shows a difference between the measure in question and a result evaluated at the present time. Or it is the ability of the test to estimate present

performance the degree to which a test correlates with an external criterion that is measured currently. As Demisse (1995) expressed, concurrent validity has such ability to show the validity of the new test. He noted that if any two tests relate to each other highly, then the new test will be valid and provide the first one extremely dependable of the test taker's skill.

Underhill (1991) noted that in order to achieve a good concurrent validity between two tests, the correlations constant need to be 0.9 or further which reveals that the same thing has been measured by the tests. On the other hand, if the correlation constant or coefficient is 0.4 or less, then both tests have lower concurrent validity. So, both figures will lose the huge majority of concurrent validity between them and no particular result can be shown from them. Essentially, the value of correlation coefficient is between +1 and -1. If we think logically, the concurrent validity coefficients become higher when two tests are used at the same time and both of them are from the same skill. For instance, two verbal test scores relate more highly than one verbal test score with multiple choice grammar test score (Bandoro, 2014).

- **Predictive validity**

Fulcher and Davidson (2007) defined predictive validity as the ability of the test in order to predict future performance which means the outcome will be evaluated at a future time, for example, academic success. Also, it's described as a vital part of the case of placement tests, aptitude tests and achievement tests.

There is similarity between predictive and concurrent validity because both types are participating in mathematical formulations for the sake of calculating validity coefficients. But if we concentrate on time gap between both test types, then we can

observe that time gap of the concurrent validity is broader than the time gap of predictive validity.

2.7.5.4 Consequential Validity

Consequential validity refers to the positive or negative social consequences of a particular test. For example, the consequential validity of standardized tests includes many positive attributes, including improved student learning and motivation and ensuring that all students to have access to equal classroom content (Messick, 1989).

Bachman and Palmer (1996), Mckay (2000), Davies (2003) and Choi (2008) employed the term impact as it relates to the consequential validity possibly more broadly inclosing many results of the assessment, beforehand and afterward the test administration.

Consequential validity has two main levels; the first one is called macro and the second one is micro level. Micro has an impact on individual examinees, while both society and educational system have been influenced by macro level. Moreover, micro level creates washback in the classroom (Bachman & Palmer, 1996).

2.7.5.5 Content validity

Content validity is defined by most of the testing experts in various ways. Generally, they focused on the same definition centrally but in different shapes.

For instance, Harrison (1983), Alderson et al. (1995) believed that content validity is the ability of a test which represents the whole contents of a specific area (syllabus) in a related method. Similarly, Heaton (1988) defined content validity and stated that a test ought to be constructed well such that it represents all the samples that have been taken through the course and then the testers should give a clear relation between the

course objectives and the test items. Also, Bachman (1990) said that if any attempt in the content of the test sufficiently represents the behavioral domain in question it is called content validity. Additionally, Weir (1990) mentioned content validity is the sample of the exam extensively as likely related, both communicative and critical items from the syllabus due to having a positive wash-back effect on teaching.

Besides, Henning (1987) and Hughes, (1989) pointed out a test has content validity if the content of a test is well represented to the content of the course in terms of its structures (content areas) and skills. If you want to test a test in terms of grammar, for instance, present continuous tense, you should not include any item related to past continuous in this test, also the test ought to claim that the knowledge is totally about present continuous tense. Likewise, in a writing test, it is not possible to ask about listening item to the examinees. A writing test instead of writing doesn't show the test has content validity. The test will be accepted to have content validity to represent those things that have been taught throughout the course (Brown, 1996).

Based on all the definitions above, according to the scholars' point of view, all the notions are almost similar in the way when they have defined content validity. They believed that the given test should be the reflection of the contents that are available in the syllabus by illustrating demonstrative examples in a relational manner. According to Anastasi (1982), a given test needs to cover representative examples of the contents of language from a syllabus that can be determined by the systematized examination of the content of the test.

Underhill (1991) said that an ordinary way of introducing the content validation is the contents of the test, it needs to be analyzed and the statements should be compared to

what the content has to be. The statements can be syllabus or a particular domain. Thus, in order to analyze the content validity of tests, the test samples and course book contents ought to be measured in terms of regularities of practice items by a comparison with the course book and items of the test in sample tests. In order to facilitate the test content analysis, it is worthwhile to utilize the table of test content specification, which is advisable (Hughes, 1989).

So, content validity is a feature of the test, but not the scores. Logical analysis of the test's content is made to see whether the content is a representative sample of the field to be tested which means the test samples represent the content of the course book that the test takers have studied in their authentic lives (Fulcher & Davidson, 2007). For example; assess someone's ability to speak a second language through conversational.

Setting: to ask students answer multiple choice questions.
Negotiate with some category of authentic context.

Content-related evidence validity is established by the followings:

1. Classroom objectives and lesson objectives should be represented in the form of test specifications.
2. The performance of test-takers should reflect the classroom objectives.

Brown (2004) showed another reason to understand clearly in content validity such as: Direct testing and indirect testing: we can achieve content validity in assessment through testing performance directly such as oral production of syllable stress. On the other hand, indirect test marks the stressed syllables in a list of written test.

41

2.7.5.5.1 The Importance of Content Validity in Language Learning

Although all the features of validity are important in the process of teaching-learning, the most proportional one will be named as content validity since it is a resource to ask the attainment of objects of the syllabus of any content (Hughes, 1989). If a test has content validity it encourages the students to study hard in any content that they have studied during their courses. In contrary, students do not study or avoid from those language areas which are not included in the test. They only concentrate on those language areas that they have to be tested.

As it is known, the expectation of testing process will make wash back effect on learning; students prepare themselves in order to take their tests, arrange information in memory in the way how they are going to take their test or how they are going to be tested. Hence, the quality and quantity of learning is affected by evaluation. Also, it is crucial to test in the realms of the process of learning and the learning results (kohonom, 1999) as cited in (Asmerah, 2008).

Similarly, if the tests are well-emphasized, learners might possibly do well in language areas and skills; in this way, during the study and performance learners focus on language areas to which more highlighting is made during the examination (Weir, 1993). Weir indicated that by looking at all contents of syllabus equivalently during the exam time, it helps to learn the language in an easy way. Finally, scores will be achieved by tests when participating content validity can be helpful to approximate learners' real practice level. That is, they provide assurance to draw satisfactory statements about students' proficiency standing in a language at a specific grade level.

In brief, as Hughes (2003) and Mousavi (2009) mentioned, if the test samples are related to the content areas of a given syllabus, a positive influence can be seen on the process of teaching-learning. It is very significant that the test items reflect widely to the contents of the language areas and the four skills.

In conclusion, they believed that decision is the best way to make about learners' activity level the syllabus can be more satisfactory. Decisions, referred to the scores attained by the tests that are poor or weak in content validity, are probable to say this test will not be acceptable. The quality and the quantity of learning will be increased when tests have appropriate content validity. Therefore, it includes valid and reliable knowledge about learners' activities and then a new form of teaching procedure is created.

2.7.5.5.2 Collecting Evidence for Content Validity of English Language Test

There are different research methods that are evaluating content validity such as mixed method, quantitative method and qualitative method. Firstly, a mixed method contains both quantitative and qualitative data, methods, methodologies, and paradigms in a research study. Secondly, quantitative research is a method the researcher works on observable phenomena by using statistical or mathematical techniques. In this method, the research utilizes instruments like CVI, chi-square and so on. The last one is qualitative research which investigates totally into particular experiences, describing and exploring meaning by the test, narrative, etc. In another way, in the qualitative method, most of the researchers are against the quantitative method because they believed that the expert judgment is the best way to evaluate qualitative method in content validity, for instance, Ghiselli (1964) stated that content validity depends on subjective or professional judgment. Moreover, Yaghmale (2003) pointed out the best

way to examine content validity in the judgment stage that it needs expert judgment to determine the extent to which the scale was designed to measure a trait of interest.

2.7.5.5.3 Guidelines to Establish Content Validity

As it was discussed before, the crucial part of any test is content validity. Most of the researchers believed that if a test has content validity it supports learning a language and different subjects in an easy way. Thus, test constructors ought to pay attention while they are preparing tests. Some precious recommendations have been proposed to the test writers in order to create content validity (Anastasi, 1982). These recommendations are as follows:

- The content area to be tested, it is necessary to be logically examined to verify that all major aspects are related to the test items and in the right extent.

- The area under consideration ought to be completely defined before, instead of being described after the test has been ready.

- Content validity relies upon the relatedness of person's test reply to the behavior domain under consideration, as opposed to rather than on the clear relation of the item content.

2.8 Related Studies in Different Contexts

Four studies have been carried out in the realm of content validity to evaluate English language tests according to their contexts. The whole research discussed the significance of content validity. The first study was conducted by Bandoro (2014) in Indonesia under the title of content validity of English language national exam. The research was designed to analyze the English language test and to what extent the test samples are related to its content in terms of language skills. The result showed that the English national exam did not represent 100% of the contents which had been taught by using the language skills.

In addition, another study was conducted at Awassa College of Health Science: by Asmerah (2008) in Ethiopia, in the title of an assessment of the content validity of English Language tests and the research discussed the correlation between the English language test and the content of the 9th and 10th grade course book in terms of content areas and language skills. The result showed that the test samples didn't sufficiently reflect the course book's coverage. Particularly, grammar, pronunciation and listening were ignored in using the test papers. Thus, a poor content validity has been represented in this study.

Moreover, in another study which was conducted by Yibrah et al. (2014) in the school of foreign language at Haramay University in Ethiopian secondary schools attempted to find out the relation between the content of the standardized achievement test (SAT) and the syllabi in terms of language areas and four skills. The result exposed that a weak relationship was found between the test sample's items and the content of the textbook. So, the study discovered that the secondary exam papers violated the content validity.

Similarly, another research was carried out at Haramaya University, in Ethiopia by Yibrah (2017) which was examined to demonstrate to what extent the English test items represented the content coverage and the objective that was discussed by the textbook. The research concluded that the test papers were not adequately represented the whole sections of the language items. Especially, in most cases, the exam papers emphasized on using grammar and reading while the listening skills were not mentioned at all.

As to the knowledge of the researcher, no studies have been done at Erbil governorate-general educational directorate till now about this issue. This is the reason behind choosing validation of English language tests of seventh grade at basic schools in Erbil governorate.

2.9 Summary

Chapter two concerned with the review of related literature and the basic goal of this chapter gave information about the aim of testing and its principles such as reliability, validity, practicality, washback and authenticity in the process of teaching and learning. Also, assessment and its components have been discussed. Validity in general and content validity particularly are discussed in this chapter because this study aimed to evaluate tests in terms of its content validity. Moreover, different test samples of seventh grade at Erbil governorate basic schools have been used in order to analyze the items according to the needs of the study. In addition, the literature review gave evidence for using qualitative research in content validity.

Finally, conducting this kind of research in the area of content validity was very significant in researcher's context because no other study has been done before in the researcher's context. Chapter three deals with the methodology used in the writer's study.

Chapter 3

METHODOLOGY

This chapter presents the whole description and the discussion of the research methodology that is utilized in this research. The description contains research design, the context of the study, data collection instruments, and data collection procedures.

3.1 Research Design

This study is a case study, and the aim of this research is to evaluate tests in terms of content validity. This study was planned to analyze the content of the final English test samples of seventh grade at Erbil governorate basic schools in the academic year of (2016-2017). The data was collected by analyzing the materials in the process of teaching and the exam samples (summative). Later, the materials will be correlated with the test samples in terms of their contents. In order to achieve the aim mentioned above, the study used a qualitative method. For the qualitative data, descriptive statistics were used to analyze the test sample items.

Content analysis is proposed to any person who desires to place an inquiry to communications in order to achieve information that will allow him/her to reach the convinced outcome (Carney, 1972) as cited in (Widyanata, 2004). It aims at comparing the data against some models, standard or theory, such as drawing its result. As it was mentioned above, this study describes the content validity of the final exams of seventh grades in the academic year of (2016-2017) at Erbil governorate basic schools in terms of its content area and skills.

Thus, the data is processed by SPSS program and the mean part of descriptive statistics will be used for analyzing the data. According to Fraenkel et al. (2011) "the major advantage of descriptive statistics is that they permit researchers to describe the information contained in many scores with just a few indices, such as the mean and median" (p. 187).

3.2 The Context of the Study

This research was conducted in general educational directorates of Erbil governorate seventh grade basic schools. The test materials for seventh grade at basic schools were collected from the general educational directorates which were provided in the center of Erbil and surrounded educational areas of Erbil governorate in Kurdistan region-Iraq. The total basic schools located in the center of Erbil, Erbil surrounded areas and Erbil countryside's. The study was conducted in September 2017.

In this regard, both public and private schools were used in this study. Additionally, both systems are under the supervision of the ministry of education in Kurdistan Region-Iraq. In contrast, both of them follow two different curriculums and use different test items as well. Erbil general directorate educations are divided into eleven education directorates which are geographically divided into Erbil center and the surrounding areas of Erbil city.

The ministry of education supervises the private and public schools similarly while the private sectors have their own curriculums totally different from the public sectors because students will learn the whole lessons in English except Kurdish language, whereas the public schools have only one English lesson. Furthermore, an obvious distinction will be seen in both sectors test samples.

3.3 Data Collection Instrument

In this research data was collected through the tests collected for the purpose of this study. They were also analyzed and interpreted to analyze the content validity. Moreover, the researcher evaluated the English language test items of Erbil governorate basic schools-7th grade by comparing their language areas such as vocabulary, grammar and pronunciation, and the four language skills like writing, reading, listening and speaking. As it was mentioned in the literature review, the researcher has permission as an expert judgment to evaluate content validity. For this reason, Sartori and Pasini (2007) stated that the essential method to decide whether a test has content validity is expert judgment. In this realm, it will be known as a qualitative part of content validity.

3.3.1 A Brief Summary of the Course Book

Sunrise is a localized English course book; it is used by the ministry of education in the recent years in Kurdistan region. This system is a new system of learning English which includes basic education which started from first till ninth grade and preparatory schools. Sunrise course book is designed by Macmillan publishers; it is designed for all the school levels which are started from level one and will be ended in level twelve in Kurdistan region public and private schools. These levels cover the students who are studying in the basic and preparatory levels.

The earliest sequence of sunrise was published in 1999. Also, at the same year Katherine Mendelsohn, Judy Garton- Sprenger, Nick Mclver, and Philip Prowse wrote the course pack test, and the book was reviewed in 2006. Three types of course book were utilized for each level, for instance, the book that is used by the teachers, the student's book and the last one called activity book. Both teachers and students can

use the students and the activity books but the teacher's book is only used by the teachers.

According to Mohammed (2015) various approaches were employed when the course pack was designed, for example, PPP (presentation, practice, and production) which is emphasized on group and pair work activities. The language area and four skills were mentioned as well. According to Mohammed, it was designed for the learner-centered approach.

3.3.2 Sunrise 7

As it is shown at the beginning of the course book, the first unit introduced the welcome unit, in the middle seven units were demonstrated, and the last one is the farewell unit. The word list and the grammar were translated into the Kurdish language.

The welcome unit which started at the very beginning of the book includes eight lessons and it is used for the beginner level in order to make easier for the students by using pictures, colors, and letters such as hello, how are you, what is your name? Etc. Hence, this could be exposed as a reviewing part that had been studied in the preceding years. The remaining seven units consist of structure skills of sounds and vocabulary see appendix (I). At the end of each unit, the review part will be mentioned. The whole seven units were intensively focusing on the language areas such as grammar, vocabulary, and pronunciation, and language skills as reading, writing, listening and speaking. The farewell unit is the smallest unit in this course book, it is brief and no lessons would be included in this chapter. It accepts as a distinct lesson by itself. This part contains the four language skills and listening songs through practicing in daily

class for the students to be familiar with listening and then all the songs are about friends.

3.4 Data Collection Procedure

The data was collected from Erbil general educational directorate and it is divided into 11 educational directorates and the test samples of the seventh grade used in this research in the academic year of 2016-2017. The collection of the test samples started from the beginning of August and ended in the middle of September.

For getting permission, the researcher had to write a petition to the head of Erbil general educational directorate in order to collect the test samples from the whole educational directorates, after evaluating the request by the head of Erbil general educational directorate. The researcher took permission to start collecting the data in 11 educational directorates see appendix (J). The next step was to submit the permission letter to the ethical committee of the faculty of education at Eastern Mediterranean University. After one and half month, the writer got approval in order to start collecting the data in his country appendix (K).

At the very first step of collecting the test samples, the researcher went to Erbil general educational directorate-supervision department to collect the test samples of 7th grade of the whole educational directorates as a PDF sample but there weren't any PDF tests in Erbil general educational directorate. So, the writer had to work on another alternative which was visiting the whole educational department and also getting approval from the heads of the whole educational directorates and then he had to visit every supervision part in order to achieve the test samples there because after every semester the schools have to send their test samples to the supervision department in

order to evaluate the test samples by the supervisors. Moreover, the writer visited all educational directorates from the center to the surrounding areas of Erbil in order to attain an appropriate number of test samples for collecting the data.

The researcher, in his first step, started from the center of Erbil and then the surrounding areas. During collecting the test samples, the researcher faced lots of problems, for example, lack of having sufficient test samples in the supervision departments because most of the schools didn't send their test samples. In addition, some of them had an excuse and they expressed "since we don't have English supervisors for that reason, we didn't ask the schools to send their test samples". Therefore, the researcher had to choose another method which was achieving test samples through the English language teachers who are related to the Erbil general directorate schools.

3.5 Data Analysis

In order to analyze the qualitative data, the writer used descriptive statistics in SPSS program by entering the achieved data to the program and the results were listed by displaying from the tables. The researcher in this study used Likert scales to recognize how to employ the level of each test items (much= 2, little= 1, no= 0) and the first part of descriptive statistics (mean) is used as well. To conclude, after analyzing the data, the writer described and discussed the findings of the seventh-grade test samples and then evaluated them in terms of their content validity.

3.6 Summary

In this chapter, the methodology of the study has been discussed which is focused on the research design, the context of the study, data collection, a brief summary about

the seventh-grade course book and the analyzation of the data as well. The next chapter will analyze the data and the discussion of the research.

Chapter 4

DATA ANALYSIS

4.1 Introduction

This chapter consists of the discussion of two tables and the structure of test items. The first one concerns analyzing the language areas, which includes three parts such as grammar, vocabulary and pronunciation. The second table consists of analyzing the four language skills such as writing, reading, speaking and listening. Each test is divided into seven parts according to its content by using both language areas and the four language skills. The three Likert scales are used to recognize how to employ the level of each test item (much= 2, little= 1, no= 0) and the first part of descriptive statistics (mean) is used. Furthermore, 160 of the final English test samples of seventh grades at Erbil governorate basic schools are evaluated. Moreover, the collected data is analyzed by using descriptive statistics in SPSS program.

The first table analyses the language areas of the test in terms of its content and demonstrates the relationship between the test items and content validity. The language area is divided into three parts and the three Likert scales for analyzing the data with one part of descriptive statistics which is called as mean is used. After analyzing the table, a sample of each part such as grammar, vocabulary and pronunciation is illustrated.

The second table is distributed into four parts which includes writing, reading, speaking and listening. The same Likert scales and descriptive statistics are used in both tables.

The structure of the test items of the 2016 English final term of the seventh-grade exam includes 5 grammar samples, 4 vocabulary and pronunciation test samples respectively, which are divided into three language area parts. The writing and reading part of the four language skills regards five writing test samples and reading test items consisting of 5 test samples.

Hence, the structure of the test is summarized in the following table:

4.2 Language Areas

Table 2: Language Areas

Language areas	Much	Little	No	D.S Mean
Grammar	96	3	1	1.96
Vocabulary	84	15	1	1.84
Pronunciation	22	46	32	.90

D.S= descriptive statistics

Based on the analysis as it is shown in Table 1 according to 160 English language exams of the final term, in the academic year of 2016 in Erbil governorate seventh grade basic schools, the first table concerns with the language areas that are classified into three sections such as grammar, vocabulary and pronunciation. In addition, three Likert scale types are used to show the percentage of the data and the (mean) of descriptive statistics, as well.

As it is indicated in the table, the first part starts with grammar, the examiners concentrated on grammar in the content of the tests rather than other items of the language areas in their exams. Regarding the data, grammar has taken the highest share and covers (96%) as it is shown the utmost part of the exams are related to the grammar whereas the other parts constituted (3%) and (1%) respectively. For the mean part of descriptive statistics, grammar comprises (1.96) out of (2) which indicates that grammar has the greatest role among the other language areas in consuming the exams.

4.2.1 Grammar Samples

The grammar section contains sentences within a blank and four multiple choices given in order to choose the correct one to complete the sentence. Hama (2015) stated that "in most of the test items the examiners are concentrated on the multiple choices by using various options in different parts of the test."

Grammar:

Instruction: Choose the correct answer.

1. Silvia is ------ a video at the moment. (watch, watching, watches, watched)

2. Anna ------ to the swimming pool every day. (goes- go- going- to go)

3. ------ there any windows? (am – is – are -be)

4. They ------- at home yesterday. (is – are –was - were)

5. Can you see the panda? ------ (yes, I see – yes, I do –yes, I can)

The second section of the table is vocabulary, based on the collected data, vocabulary takes the second place which the exams use and it comprises (84%) in the overall test but the little one includes (15%) as it is indicated while only (1%) doesn't use vocabulary in the test samples which indicates that the percentage of not using

vocabulary is one percent. According to the mean in descriptive statistics, vocabulary takes the second part after grammar and it has the share (1.84) out of (2).

4.2.2 Vocabulary Samples

This part illustrates four options on the cover of multiple choice and asks to find the odd words.

Vocabulary:

Instruction: Find the odd words.

(May-August-September-Tuesday)

(Apple-Sandwich-Orange-Banana)

(Coat-Grey-Brown-Blue)

(Large-Medium-Small-New)

The last category of language area is pronunciation which is less concentrated in the test exams and the percentage contains (22%) as it is shown in the table, however, the little-considered part increases to (46%), on the other hand, (32%) not includes pronunciation in the exams. As it is clear in the exams, the examiners concentrate less on pronunciation than grammar and vocabulary so pronunciation becomes disregarded part in using the test if we compare with grammar and vocabulary. The mean part coverage is only (.90) out of (2).

4.2.3 Pronunciation Sample

For pronunciation, the examiner asks the examinees to match the two words that have the same rhyme.

Pronunciation:

Instruction: Match the rhyming words.

A	B
Two	Pen
Eleven	Band
Ten	You
Hand	Seven

4.3 Language Skills

Table 3: Language Skills

Language skills	Much	Little	No	mean
Writing	77.5	2.5	20	1.58
Reading	75	20	5	1.10
Speaking	0	0	100	0
Listening	0	0	100	0

Table 3 indicates the four English language skills that are divided into four sections that are writing, reading, speaking and listening. It is clear that the test developers in their exam papers focused on writing first in the final terms. As it is shown in the table, in the writing part covers the highest share, which is (77.5) whereas the little consideration constitutes (2.5%) in the content of the tests, moreover, (20%) of the test samples don't use writing in the whole tests. Furthermore, the mean of the writing part contains (1.58) out of (2), this skill more emphasized than the rest of the other skills.

4.3.1 Writing Sample

This part is designed in two different ways, the first way asks five or six questions and the examiners have to answer the questions in a paragraph. The second way is asking a question, for example, write a paragraph about your best friend.

Writing:

Instruction: Answer the following questions in a paragraph.

1. What is your name?

2. Where are you from?

3. How old are you?

4. What is your favorite sport?

5. What language do you speak?

(Or write a paragraph about your best friend)

Reading becomes the second part of the language skills in Table 2, it is clear that there is no big difference between reading and writing skills as it can be understood from the content of the test. Moreover, writing and reading comprise (77.5%) and (75%) in test samples respectively. In contrast, (20%) comprises less reading items in the test papers. Conversely, (5%) doesn't include reading in the exams which takes the lowest coverage from the content of the test. If we look at the mean, it is (1.10) out of (2) which is a little bit rather than half of the degree.

4.3.2 Reading Samples

The reading part gives a passage to the examiners and asks to read the passage, and then to write the true and false.

Reading:

Read the following passage carefully, then write (true) or (false).

Aso comes from Erbil. His birthday is on 14th of April. He has got two brothers and one sister. He is one of the best students in the class. He likes all his friends. He has short black hair and blue eyes. His favorite clothes are shirt and jacket. He likes swimming but he hates tennis.

Instruction: Read the sentences and mark them [T] for true and [F] for false.

1. He likes swimming.
2. He has got long black hair and brown eyes.
3. His favorite clothes are T-shirt and jeans.
4. He has one brother and sister.
5. His birthday is on 14th of April.

Based on speaking and listening skills, the table shows from the overall test that the test developers do not apply speaking and listening skills in their tests. Both skills are ignored totally and they take the lowest percentage which is (0%) for speaking and listening items respectively. Furthermore, the data reveals that the examiners do not use speaking and listening in the summative tests. Similarly, the mean part of descriptive statistics for the former skills comprised of (0) out of (2) as it exposes on the table above.

4.4 Summary

Chapter four analyzed the data of both the language areas and the four language skills through two different tables and then the two tables were discussed. Finally, chapter

five shows the discussion of the result, conclusion, implication, limitation and the recommendation related to this study.

Chapter 5

RESULTS AND DISCUSSION

5.1 Introduction

In this chapter the findings will be discussed, the research questions will be answered and the relationship with the correlated studies in the literature review will be set. These are done through the conclusion and practical implications as well. Lastly, this study discusses some limitations, and several suggestions are proposed for further studies.

5.2 Discussion of the Results

This study attempts to evaluate tests of seventh grade in Erbil governorate-general directorate in terms of their content validity. The study examines how the test samples have reflected the content of the materials that have been taught through language areas and the four language skills. In this study, a qualitative research is used because it is analyzing the test materials which mean to test the test by comparing testing principles.

In order to find out the relationship between the test items and contents in the realm of the language areas and skills, the following research questions are asked:

5.2.1 Research question 1: To What Extent do the English Language Tests Used at Erbil Basic Schools Reflect the Content of the Materials been Taught in Terms of Language Areas? (Grammar, Vocabulary and Pronunciation).

The findings of the study revealed that the highest coverage of utilizing the test samples in their language areas goes to the grammar, as it is shown, the practitioners focused on the grammar as a first item, vocabulary as a second item and pronunciation takes the third place in the content of the exams. This result contradicts the Asmere's (2008) finding, in which the examiners on the test papers emphasized more on vocabulary than grammar. On the other hand, in Yibrah et al. (2012) finding, grammar had the highest percentage among the language areas part in using on the exam papers. Based on the vocabulary, it became the second item after grammar in using the exam samples.

Equally, this finding is similar to Yibrah's (2017) study, the grammar dominated the exam papers as the highest share but vocabulary items represented the smallest coverage in the final exam samples.

Vocabulary has become the second part of using the test samples after grammar in this research. As it is discovered, the use of language areas on the exams are not similar according to the test items, the reason refers to the instructors when they focus more on a particular language area than others in their tests. On the other hand, in Asmere's (2008) finding, vocabulary took the highest share in using the exam papers among other parts of the language areas.

In contrast, the smallest test part in language areas has referred to the pronunciation. It is found that pronunciation has become an ignored part among the mentioned parts

above. It is clear that pronunciation is not seen as a major part of the test samples since only small items of the tests focused on the pronunciation. Similarly, in Asmare's (2008) finding, the examiners less concentrated on pronunciation in the exams if we compare with the rest language areas. Additionally, in Yibrah et al. (2012) and Yibrah's (2017) findings, pronunciation is ignored totally by the test developers in utilizing on the test papers which is a major part of language areas. Therefore, a big difference has been displayed in employing the number of items by the examiners in designing the exams among the language areas in the final exams as it was discussed above.

In summary, the majority of English language tests in seventh grade at basic schools in Erbil governorate emphasized on grammar as a first item in using on the test papers among the rest language areas. This evidence demonstrated that grammar has become a significant part than the other two language areas to use in the summative tests at the end of the semester by the examiners. The situation of vocabulary came after grammar with a small distinction which was detected that both parts were almost closed to each other. On the contrary, the pronunciation part is ignored by the instructors as it is not adequately focused in the tests and a big difference can be shown if we compare the using of pronunciation with grammar and vocabulary according to the discussion part what has been discussed above.

5.2.2 Research question 2: To What Extent do the English Language Tests Used at Erbil Basic Schools Reflect the Content of the Materials been Taught in Terms of Language Skills? (Reading, Writing, Speaking and Listening).

Based on the results that obtained in this research in using the materials in order to reflect the language skills in the final exam of seventh grade, the test developers focused first on writing, reading stayed at the second stage and then speaking and

listening were not mentioned on the exam papers respectively, which means the teachers didn't use speaking and listening items in the final exam at all. So, it is clear the testers emphasized more on writing as a first material in employing the exams. However, in Bandoro's (2014) finding, the writing was the smallest item in participating on the test papers and the teachers concentrated on the rest skills rather than writing; and then, the rest skills will be demonstrated in the following paragraphs.

Reading came after writing while the distinction between writing with reading in using the examinations was not much as it was shown in Table 2 in Chapter 4. Contrarily, in Yibrah et al. (2012) finding, reading had the highest rate in using the exam papers while writing came after reading and then speaking and listening followed each other respectively. Similarly, in Yibrah's (2017) study, reading was used in the exam papers as a first item and both writing and speaking followed each other in a series. In contrast with the former study, listening skill wasn't mentioned in the exams in this study.

As it was explained, in Erbil governorate basic schools writing and reading items were almost close to each other in terms of usage number. If we focus on Bandoro's finding, we will find a contradictory result because the reading skill was taken as the highest coverage by utilizing the number of its items in the exams rather than listening and writing.

Listening and speaking were ignored completely in the final exams of seventh grade, the reason behind this relates to the testers of the basic schools because they do not see both skills as a crucial part of the test items. While both skills were discussed in detail in every chapter of the seventh-grade course book and students and teachers should practice those skills in the class on daily basis. In contrast, in Bandoro's finding,

listening became the first part in using its items on the test samples taking the maximum coverage in employing the exams by the teachers. For the speaking skill, the situation was different in Bandoro's finding because he didn't mention about speaking skills in his research at all.

To sum up, the results of using the four language skills of the final exam of seventh grade showed that the test developers emphasized more on writing as an important part in contributing on the exam papers. The second skill which was also used in the tests is reading but with a small variety of using its items with writing as mentioned before. However, the listening and speaking skills were not used in the exams entirely, which revealed that both skills were ignored by the examiners to contribute to the final exam.

5.3 Conclusion

Overall, the finding of this research revealed that language area parts of the final exams of seventh grade such as grammar and vocabulary mostly reflected the content that is utilized in everyday classroom teaching while pronunciation rarely reflected the content validity of the test. Furthermore, the four language skills such as writing and reading reflected adequately the content that was employed in the classroom activates whereas the speaking and listening skills were totally ignored in the test samples.

The reason behind that may refer to the teachers or testers because they didn't use the whole items on the test papers correspondingly. In addition, the education system in KRG is also responsible because the test experts of the Ministry of Education will guide the instructors about how to design the test samples. Furthermore, lack of training course to the examiners and teaching them by the test experts on how their tests reflect the content that they have taught in the class time.

If we take a look at the language area, the test specialists emphasized on grammar rather than other parts. Grammar took the highest share of using the exam samples but for vocabulary item, the number of the item decreased in using the test. At that time, it was discovered that the test item samples fluctuated from one part to another. For pronunciation as the last part of language area, this part had the smallest coverage in the exams and it was less emphasized among the other language areas. These verities in using the items show that the final tests of Erbil governorate basic schools are violating the content validity in language testing. When one part is ignored and the others are cared to be the major part of the test, then the test loses its validity according to the validity principles.

Furthermore, the finding of language skills showed different results. At the very beginning, the writing skill became the first skill that concentrated more than the other three remaining language skills which are reading, listening and speaking. For the reading part, a small distinction were found between writing and reading in taking part of the exam papers. On the other hand, listening and speaking skills were forgotten completely in the whole exams. As it was exposed, listening and speaking were not accepted to use in in the final examination.

Finally, the result revealed that among four language skills in using the final seventh-grade test writing and reading became a major part of the test papers whereas listening and speaking were ignored totally by the test makers. Due to these shortages, English language final tests of the seventh grade at basic schools in Erbil governorate are found to be weak in terms of content validity as it was concluded, because they did not reflect all contents in the language areas equivalently and they avoided important language skills and knowledge from the content of the test as well.

5.4 The Pedagogical Implications of the Study

The findings of this study may provide useful feedback to the English language supervisors and test designers about the effectiveness of tests that are used by teachers. In addition, the findings may help to make possible changes in test format given by the Ministry of Education and supervised by the general educational directorate in Erbil basic schools Kurdistan Region-Iraq.

5.5 Limitations of the Study

It is worthy to say there are some limitations in this study:

First limitation is the lack of participants in order to interview them in the capacity of specialists in language testing such as language supervisors and test designers because it is their responsibility to design and prepare the test. Additionally, in order to enrich the result, it was very crucial to have both characters' opinions.

Secondly, the schools used in collecting the data are related only to the seventh grade at basic schools in Erbil governorate. This shows that the other cities' basic schools are not used in this research and it can be accepted as a limitation. The finding emphasized only on one general educational directorate seventh grade basic schools, but it is not generalized to all the basic schools-7th grade in Kurdistan Region-Iraq.

Finally, lack of achieving lots of test samples is also another limitation because only 160 test samples were obtained from 11 educational directorates.

5.6 Suggestions for Further Research

Based on what has been discussed in the research, the researcher would like to present some suggestions for the next research in the same field by considering the limitations as in the following:

In this study, only test samples were used for collecting the data, neither interview nor questionnaire was used which means that this research doesn't have any participant. Using participants particularly making an interview with the language supervisor and test designer may provide more precise results about the research topic. This research was carried out in the seventh grade at basic schools in Erbil governorate. The finding can be limited to only one general educational directorate and 7th grade as well. It is very crucial to use more than one general education or the whole general education in Kurdistan region as well as 8th and 9th grade test samples. This will help other researchers that they will decide to work on the content validity in the future.

Another suggestion for those who work in the same realm in the future is gathering extra test samples for the data collection. Furthermore, it is noteworthy that both formative and summative tests in one year of study will participate for the sake of knowing well the content validity of the tests.

5.7 Summary

In chapter five the finding of the study, the answers to the research questions and the related studies were discussed, in addition to the result of the research and practical implication. Finally, some limitations and suggestions were proposed for future studies.

REFERENCES

Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Munich: Ernst Klett Sprachen.

Amin, Y. M. (2017). English language teaching methods reforms in English curriculum in Iraq; An overview, *JUHD, 3*(3), 578-583.

Anastasi, A. (1982). *Psychological testing*. New York: Macmillan.

Angeles, P. A. (1981). *A dictionary of philosophy*. New York: Barnes & Noble.

Asmare, M. (2008). *An assessment of the content validity of English language tests*: The case of Awassa College of Health Sciences (master's thesis). Addis Ababa University, Ethiopia.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. London: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.

Bandoro, J. C. (2014). *The content validity of the 2013 English national exam* (master's thesis), Widya Mandala Catholic University, Surabaya, Indonesia.

Brown, D. H., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices.* New York: Pearson Education.

Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. New York: Longman.

Burton, S. J., Sudweeks, R. R., Merrill, P. F., & Wood, B. (1991). *How to prepare better multiple-choice items*: *Guidelines for university faculty* [Booklet]. Utah: Brigham Young University.

Cheng, L. (1997). How does washback influence teaching? Implications for Hong Kong. *Language Education, 11*(1), 38–54.

Clapham, C. (2000). Assessment and testing. *Annual Review of Applied Linguistics, 20,* 147–161.

Clegg, V. L., & Cashin W. E. (1986). *Improving multiple-choice tests*. Retrieved from https://www.ideaedu.org/Portals/0/Uploads/Documents/IDEA%20Papers/IDEA%20Papers/Idea_Paper_16.pdf

Davies, A. (1999). *Dictionary of language testing*. Cambridge: Cambridge University Press.

Davies, A., & Elder, C. (2005). Validity and validation in language testing. In E. Hinkel(ed.), *Handbook of research in second language teaching and learning* (pp. 795-813). New York: Routledge

Dejene, L. (1994). *Testing and its practical application: Testing reading in focus* (Unpublished Seminar Paper). AAU: Department of Foreign Language and Literature.

Demisse, T. (1995). *The construction and validation of a test in English for* Addis Ababa (Unpublished Doctoral Dissertation). Addis Ababa, Addis Ababa University.

Douglas, B. H. (2004). *Language assessment principle and classroom practices*. NY: Pearson Education.

Duran, R. P., Canale, M., Penfield, J., Stansfield, C. W., & Gasparro, J. E. (1985). TOEFL from a communicative viewpoint on language proficiency. *ETS Research Report Series*, *1985*(1), 1-46.

Frisbie, D. A., & Ebel, R. L. (1991). *Essentials of educational measurement*. New Delhi: Prentice Hall College Div.

Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing, 13*(2), 208–238.

Fulcher, G. (2013). *Practical language testing*. New York: Routledge.

Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London, England & New York, NY: Routledge.

Ghiselli, E. E. (1964). *Theory of psychological measurement*. New York: McGraw-Hill.

Good, C. V. (1973). *Dictionary of education*. New York: Mc. Gram Hill.

Gronlund, N. E. (1982). *Constructing achievement tests*. New Jersey: Prentice Hall.

Hama, H. Q. (2015). Multiple-choice test item in basic schools of Qaladiza town. *Raparin University Journal, 2*(3), 41-60.

Hamavandy, M., & Kiany, G. R. (2014). A historical overview on the concept of validity in language testing. *Advances in Language and Literary Studies*, *5*(4), 86-91.

Harmer, J. (2001). *The practice of language teaching* (3rd ed.). Harlow, England: Pearson Education.

Harris D. P., (1969). *Testing English as a second language*. New York: McGraw-Hill.

Harrison, D. (1983). *Language testing handbook*. New York: Cambridge University Press.

Heaton, J. B. (1988). *Writing English language test*. New York: Longman Inc.

Henning, G. (1987). *A guide to language testing*. Cambridge, MA: Newbury House Publishers.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge, MA: Cambridge University Press.

ICDR (1999). *Teacher handbook*. Addis Ababa: Ministry of Education.

Kaplan, R. M., & Saccuzzo, D. P. (2012). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Wadsworth, Centage Learning.

Lado, R. (1961). *Language testing*. London: Longman.

Lewkowicz, J. (1997). *Investigating authenticity in language testing* (Unpublished Doctoral Dissertation). England, Lancaster University.

Liesman, M. (1991). *An analysis on the construction of the 1989-1990 English UUB for the first grade senior high school in regional north Surabaya*. Widya Mandala Catholic University, Surabaya, Indonesia.

Lynch, B. K. (2003). *Language assessment and programme evaluation*. Edinburgh, Scotland: Edinburgh University Press.

Mekonnen, Y. K., D., Devardhi, J., & Nelson, D. (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), 1-13.

Messick, S. (1989). *Educational measurement*. New York: Macmillan.

Mohammed, S. J. (2015). *An analysis of the representation of multiple intelligences theory in Sunrise 7 and teacher's perspective* (master's thesis), Near East University, Nicosia, Northern Cyprus.

Morrow, K. (1977). Authentic texts and ESP. In S. Holden (ed.), *English for specific purposes* (pp. 33-48). London: Modern English Publications.

Mousavi, S. A. (2009). *An encyclopedic dictionary of language testing* (4th ed.). Tehran: Rahnma Publication Parsons Education Limited.

Nunan, D. (1999). *Second language teaching and learning*. Boston: Heinle and Heinle Publishers.

Sartori, R., & Pasini, M. (2007). Quality and quantity in test validity: how can we be sure that psychological tests measure what they have to?. *Quality & Quantity, 41*(3), 359-374.

Thirumalai, M. S., Mallikarun, B., Mohanal, S., & Sharada, B. A. (2012). *Language in India*: Strength for today and bright hope for tomorrow, *12*(1), 432-442.

Thorndike, R. L., & Hagen, E.P. (1977). *Measurement and evaluation in psychology and education*. New York: John Wiley and Sons.

Underhill, N. (1991). *Testing spoken language*. Cambridge: Cambridge University Press.

Venkateswaran, S. (1995). *Principles of teaching English*. Delhi: Vikas Publishing Pvt. Ltd.

Wall, D. (1997). Impact and washback in language. Language testing and assessment. *The Encyclopedia of Language and Education, 7,* 291–302.

Weir, C. (1990). *Communicative language teaching*. New York: Prentice Hall.

Weir, C. J. (1993). *Understanding and developing language tests*. New York: Prentice-Hall.

Widyanata, O. T. (2004). *An analysis on reading question of senior highschool English textbook based on Bloom's taxonomy of cognitive domain.* Surabaya: Widya Mandala Catholic University, Indonesia.

Yaghmale, F. (2003). Content validity and its estimation. *Journal of Medical Education*, *3*(1), 25-27.

Yibrah, M. (2017). Assessing content validity of the EGSEC English examinations. *International Journal of Innovations in TESOL and Applied Linguistics, 3*(1), 1-66.

Yoloye, E. A. (1984). *Continuous assessment: A simple guide for teachers*. London: Cassell.

APPENDICES

Appendix A

زمانی ئینگلیزی

Class	Units	Period per a week	First Semester	Second Semester
Seventh	~7	5	1-3	4-7

دابەشکردنی نمرەکانی تاقیکردنەوەکانی پۆلەکانی بازنەی سێیەم (خولی یەکەم)

دابەشکردنی نمرەکانی تاقیکردنەوەکانی پۆلی (حەوتەم) (وەرزی یەکەم) + وەرزی دووەم

(Seventh Year) (First Semester) ــ Second

First Term			Mid Exam	
Oral & Daily activity	Written	Oral & Daily activity	Written	
	50%پرسیرەکان بەشێوازی هەڵبژاردن		50%پرسیرەکان بەشێوازی هەڵبژاردن	
Reading 3	Grammar 4	Reading 3	Grammar	8
General Questions 4	Vocabulary & Pronunciation 4	General Questions 4	Questions about reading passages	4
Activities 3	Functional Language 3	Activities 3	Vocabulary	6
	50%پرسیارەکان بەشێوازی (مقالی)		Pronunciation	6
			Functional Language	6
	Vocabulary 2		50%پرسیارەکان بەشێوازی (مقالی)	
	Spelling 2		Spelling	6
	Questions about reading passages 5		Unseen passage	10
			Composition	14
Total 10	Total 20	Total 10	Total	60

6

79

دابەشکردنی نمرەکانی تاقیکردنەوەکانی پۆلەکانی بازنەی سێیەم
(خولی دووەم)نمرەکانی دەوری دووەم بەم جۆرُە دەبێت لەبەرئەوەی
هەرنمرەی زارەکی و چالاکی رۆژانە ی خولی یەکەم ئەژماردەکرێتەوە کە 20
نمرەیە. کەواتە تەنها تاقیکردنەوەی نووسین دەکرێت کە بەم جۆرەیە:

Seventh , Eighth & Ninth Year	80 Marks
Written Exam	
50%پرسیرەکان بەشێوازی هەلبژاردن	
Grammar	10
Questions about reading passages	8
Vocabulary	8
Pronunciation	8
Functional Language	6
50%پرسیارەکان بەشێوازی (مقالی)	
Spelling	8
Unseen passage	16
Composition	16
Total	80

مامۆستای بەرێز..........کار بەم خشتەیە دەکرێت بۆ سالی (2016—2017)
تا هاتنی رێنمایی نوێ لەسەرتاقیکردنەوەکان.

لیژنەیەک لە سەرپەرشتیارانی پسپێری زمانی

ئینگلیزی لە بەرێوەبەرایەتی پرۆگرام و چاپەمەنییەکان

Appendix B

Total marks:60

Azhee Basic School for boys

Grade:Seventh   Final Examination of the second course

Date: / 5 / 2017   First trial (2016- 2017)   Time:2 hours

Q1/ Read the following passage then write (true) or (false)

Ali is a nice boy. He lives in London in England. He is 14 years old and his birthday is on the fourth august. He has one brother and one sister. His favourite subject is English.

1-His name is Ahmad.   Choose (5)

2-He is 15 years old.

3-He lives in London in England.

4-His birthday is on fourth august.

5-His favourite subject is maths.

6-He has two brother and two sister .   10 marks

Q2/ Find the odd word: Choose (5)

1-Cat , Dog , Lion , Duck .

2-morning , running , swimming , dancing .

3-Big , large , small , fast .

4-Bedroom , hall , office , clock .

5-Asia , Africa , America , Antarctica .

6-cup , plate, spoon , table .   10 marks

Q3/ Match the words that rhyme: 10 marks

(any , right , eat , egg , drink)

1-leg----------- 2-think-----------3-night-----------4-many-----------5-feet------------- .

124

81

Q4/ Put out the right choice: Choose (5)

1-I come here ----------Friday.   (in , on , at)
2- He is ----------a milk .   (drinking , drink , drank)
3-A rabbit is----------than a mouse.   (big , bigger , biggest)
4-Carla is good at ---------the guitar.   (playing , play , played)
5-She---------here last night.   (was , were , am)
6-Everyday I -------to people on the radio .   (talk , talks , talking)   10 marks

Q5/ Fill in the blanks with the missing word:

1-foot , feet   man , ---------------
2-was not , wasn't   were not , ----------------
3-tall , taller   expensive , --------------------
4-make , making   wear , -------------
5-phone , phoned   close , -------------   10 marks

Q6/ Write a paragraph on this chart: 10 marks

Name   Kawa
City   Duhok in Kurdistan
Age   14 years old
Language   Kurdish and English
Favourite sport   Football
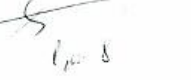
    His name is ....

    ....

    ....

Examiner teacher :   member: member :
Sarbaz fars   Barzan M. Hamad   Dilshad Ibrahim

Appendix C

Shano Basic School   Examination in English   Class: 7th
For Girls   Second Course: First Term   Time: 2 hours
     2016 : 2017   Date :^{21st} of May.

Q1/ A/ Read the following passage carefully.   (5 Marks)

One of Kamaran's feet was bigger than the other. I can never find shoes for my feet. he said to his friend Hassan. And he advises him to go to shoemaker? Hassan said. there's a good one in our village and he is quite cheap, he wrote his address to him.Kamaran went to the shoemaker a few days later, and the shoemaker made for him a pair of Shoes.Kamaran went to the shop again a week later and said to the shoemaker angrily, you area silly man, I said, Make one shoe bigger than the other "but you've made one smaller than the other.

Now write **T** if the sentences is true and **F** if the sentences false and correct the False one.

1. Kamaran's feet were same size.
2. He can't find shoes for his feet
3. The shoemaker was from Hassan's village.
4. Kamaran went back to the shoemaker two weeks later.
5. Hassanadvises himto go to shoemaker.

****

B/ Do as required.   (5 Marks)

1. Cheap → Cheaper   : exciting→ ---------------------------
2. Whose jacket is this? (Lana). It is ----------------------- (possessive S)
3. My face : I have got ------------------
4. £ 9-25 nine ------------------- twenty five.
5. Chair→ chairs   : Foot → ------------------

****

Q2/ Choose the correct answer.(Choose 10)   (20 Marks)

1. They like ----------- to music. (a-listen   b-listening   c-listened)
2. How ------------- is Sara? (a-many   b-long   c-old)
3. In Pompeii at two o'clock there ---------- a loud explosion. (a-was   b-is   c-are)
4. Its quarter to four. (a-4:15   b-5:15   c-3:45)
5. Spaghetti is ----------------. (a- accountable   b-uncountable)
6. A mouse is ------------ than a rabbit. (a- small   b- bigger   c-smaller)
7. What are these? ------------ Shirts. (a-it is   b-they are   c-there are)
8. Did Anna watch the programme? (a-yes she did   b-No she did   No she didn't)
9. Tom gets up ------- 7:00 o'clock. (a- in   b-at   c-from)
10. This is your bag. It's ------------. (a- his   b- yours   c- mine)
11. There is a door ---------- the right. (a- on   b- in   c- at)

1

Q3/A Find the odd word.   (5Marks)
1. Tiger   younger   shorter   taller.
2. Ostrich   giraffe   panda   monkey.
3. Bathroom   fridge   toilet   shower.
4. Head   hand   ticket   foot.
5. Sandwich   potato   pizza   table.
****

B/ Find the rhyming word in the box.
   Doors   Thumb   pleas   seat   grey   (5Marks)
1. Play
2. Feet
3. These
4. Yours
5. Come
****

Q4 Answer these questions. **(Choose 5)**   (10 Marks)
1. Did he like the film? X
2. Are there any chairs in the Pompeii? X
3. Does Peter like dancing? ✓
4. Can dogs run?
5. How many computers are in the space station?
6. Are there any beds in the space station?

Q5/Answer either A or B
A/ Write an email to your penfriend about what you like / don't like doing in your free time   (10 Marks)
as a paragraph
****

B / Look at the chart and write description as a paragraph.

Tiger	
Lives	Asia
Colour	Brown with black stripes
Size	up to 3 meters long
Eats	animals , fish , and birds
Can /	run fast
Can't	fly

****

Good luck

Set by: GelasA.Rasheed   Hero Hassan

Q1			
Q2			
Q3			
Q4			
Q5			

2

Appendix D

Q1/ Read the passage then answer the questions.   15 Marks

The Village Hotel is a great place for young people. You can dance at the Swing Disco. You can see films at the hotel cinema. The hotel has 600 rooms. There are two restaurants.

1. What is the place for young people? *The Village Hotel.*
2. How many restaurants are there? *There are two restaurants*
3. Where can you see films? *You can see films at the hotel ci*
4. How many rooms does the hotel have? *The hotel has 600 rooms*
5. Where can you dance? *You can dance at the Swing Disc*

Q2/ Choose the correct options. (Choose 7)   14 Marks

1. Can she walk? A. Yes, I can. **(B. Yes, she can.)** C. Yes, he can.
2. Banaz is_____ than kawa.
   (A. shorter)   B. short   C. shortest
3. I get up_____07:00 am. A. in   B. from   **(C. at)**
4. She_____ tennis. **(A. plays)** B. play   C. playing
5. There_____ a camera.   A. were   B. are   **(C. is)**
6. I hate_____ mistakes. A. make   **(B. making)** C. makes
7. Were they at home?
   A. Yes, we were.   **(B. Yes, they were.)**   C. Yes, he is.
8. Water is_____ nouns.
   A .countable   **(B. uncountable)**   C. plural

Q3/ Find the odd words of the following. (choose 5)   15 Marks

1. **(parrot)**   rabbit   lion   elephant
2. bedroom   hall   office   **(clock)**
3. **(tiger)**   bigger   shorter   taller
4. **(morning)**   running   swimming   dancing

5. pouch   (tail)   wing   hand

6. opened   liked   closed   (walk)

Q4/A/Choose the words that include the sounds.(Choose 2) 4Marks

1. /ai/   drink   sit   think   (ice)

2. /u:/   book   look   (school)   good

3. /əʊ/   now   (go)   mouth   brown

B/ write the opposites of the following. (Choose 2)   4 Marks

(early, large, false)

1. late   ×   *early*

2. small   ×   *large*

3. true   ×   *false*

Q5/ Write the missing letters of the following. (choose ⁴)   8 Marks

(g,   s,   b ,   a,   o,   zz)

1. rab_b_it   4. Sp_o_rt

2. c_a_t   5. fi_s_h

3. eg_g_s   6. pi_z_ _z_a
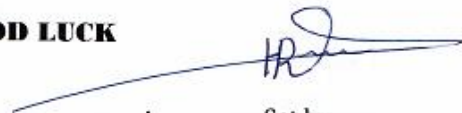
GOOD LUCK

Member   Set by

Rawand J. Jabar	Herdin I. Qadr

/19

Appendix E

Subject: English	Kurdistan Regional Government	Date: - -
Grade 7th	Brayati nwe co-ed basic school	Time: 2:00 Hours
Name:	second Course: First Attempt 2016-2017	

Q1-A/ Read the following passage carefully write (true or false) and correct the false ones.

Rawand is one of our winners from Kurdistan and he's from Hawler. He's fifteen and he lives at home with his parents and his sister, Avesta. His favourite clothes are his trainers and his tracksuit. Rawand's favourite colours are orange and blue and his favourite English word is 'yes'. Nothing makes him angry and his friends make him happy.

1- He's from Hawler.
2- His favourite clothes are shoes and his tracksuit.
3- His favourite English word is 'ok'.
4- His friend make him sad.

8 marks

B/ Read these sentences carefully and write (true or false):

1- there was a volcano called Pompeii.
2- Dolphen can make more than 20 different sounds.
3- Girrafe can bend its neck.
4- The biggest pizza in the world in South Africa.
5- Life in Pompeii started 1900 years ago.
6- Hozan is great at dancing.

6marks

Q2/ Choose the right word to fill in the gaps:

1- Anna can .... the guitar. (plays_ play_ playing)
2- .... does the film start? It starts at 10 o'clock. (where_ when_ what)
3- Banaz............ 20 km last weekend. (walk_ walked_ walks)
4- pedro doesn't............ banana. (like _ likes_ liking)
5- Are you working hard? (Yes, you are_ yes, I'm_ yes, she is)
6- What do you like doing............ Saturday? (in_ on_ at)
7- There are.............. sleeping bags. (any_ some_ a)
8- They like.............. to music. (listening_ watching_ playing)
9- I don't want these jeans, I want .... jeans. (this_ those_ that)
10- Pompeii............ quite small. (were_ was_ is)

20marks

138

1

87

Q3/Write these words in the correct column:

(enjoy_ cheaper_ over_ amazing_ office_ carry)

Noun   verb   preposition   adjective   comparative adjective

6marks

Q4-A/Do the rhyme or not?

1-touch_ much   2-goes_ does   3-bread_ eat

4-small_ wall   5-look_ moon   6-climb_ ice   6marks

B/Write these letters under the correct gapes: (n_ w_ i_ h)

1-sand....ich   2-ora.....ge   3-d....nner   4-c....eese   4marks

---

Q5/Read the profile and write paragraph about (Rick Cassidy):

Loves: eating chocolate   Hate: salad, getting up early

At weekend: likes playing basketball   good at: dancing

Bad at: answering questions

First thing in the morning: takes his dog for a walk

Last thing at night: has a shower   10marks

....

....

....

....

....

Good Luck

T. Zhwan Abubker I   T. waleed Murad

Appendix F

Q1// Read the following paragraph, Then write (True   or   False) .   10M

Banaze is a clever pupil, she is twelve years old, she always get up at seven, she washes her hand and face every day, then she bruches her teeth, after that she goes to school by bus ..

1. Banaze is a teacher .

2. she usually get up at 7 .

3. she is 12 years old .

4. she doesn't bruch her teeth .

5. she goes to school by bicycle .

Q2// Choose the correct answer.   (only 5)   10M

1. they _____got new car .   (a. has   b. have)

2. there _____ a camera . (a. are   b. is)

3. _____ he like pink . (a. does   b. do)

4. Can you sing ? yes, I _____.   (a. can   b. can't)

5. There are _____ books.   (a. any   b. some)

6. I'm watching . (a. present simple   b. present continous)

Q3// Find the odd word .   10M

1. first   twenty   eight   three
2. apple   orange   sandwich   banana
3. board   blue   black   brown
4. large   new   small   medium
5. tuesday   friday   sunday   march

Q4// Find the rhyming word .   10M

(hi , ten , fine , choose , sit)

1.excuse_____   2. nine_____   3. my _____

4. it _____   5.pen_____

89

Q5// Answer the following, (only 5)   10M
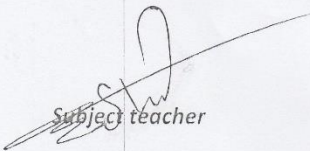
1. up ___down   , open ___

2. sock ___socks , shoe ____

3. he is ___he's , Iam ____

4. ten + one ____

5. edr ____ red   , kinp ____

6. quarter to eight ____7:45 , quarter past eight____
_____

Q6 // Answer these question .   10M

1. What is your name ? (kawa) ....

2. How old are you ? (thirteen) ....

3. Where do you live ?(Hawler) ....

4. What is your favorite sport ? (tennis) ....

5. What is your favorite color ? (blue) ....

....

Good Luck

Subject teacher   Member teacher

Shwan Muhamad .H   Hazha Dedar .R

2

Appendix G

Subject: English

Kurdistan Regional Government

Date. 14 may . 201

U ame:

Przha basic school for girls

7th grade

second course/ first attempt

time : 2:30 h

.1/ Read this passage and put the sign (T) or (F) and correct the false sentence :. 10m.

London is the capital of England. It`s In the south of the country , on the river Thames. It`s about 7000,000 eople live in London . it`s an exciting city and has more than 25000,000 visitors every year. The weather in London ; not very cold or very hot . There are lots of places to visit in London , it has a lot`s of famous building and beautif arks. In London you can eat food from many country from France , Italy , Greece , India and may more. Pizza , ebabs , fish and chips are very popular

. London is capital city of Italy.

. It`s in the north of country .

. 6000,000 people live in London.

. The weather in London is never very hot or very cold .

. London is an exciting city and has more than 25000,000 visitors every year.

.2/ Choose the correct word :. 22m

. Pompeii ---------------------- a small town .   (was   ,   were   ,   are   )

. My father was in the garden ---------------------- .   (yesterday ,   now   , usually)

. Films are more ----------------- than computer games .   (cheaps   , exciting , easy   )

. Daniel likes -------------------- .   (learn   , learns , learning)

. What would you like to ----------------- .   (eat   , eats   , eating)

. The swimming pool is ----------------- the café .   (from   , with   , near   )

. I am -------------------- at the moon and the stars now ' (looking , looked , looks)

. The game starts ----------------- 9:00 P.M .   (on   ,   at   ,   in   )

. -------------------- you watch the film yesterday?   (Do   , Does   ,   Did   )

). Are you good at ----------------- ?   (sing   , singing , sings)

L. How many bedrooms ----------------- there ?   (is   ,   am   ,   are   )

.3/ Circle the two rhyming words:. 10m.

. that   what   not

. go   do   two

. great   eat   eight

. wait   bread   hate

. know   town   noun

91

1

.4/ Put the letters in the right order :. 6m.

. ltrif

. dalsa

. illatort

. adbore

. pschi

. trwa

.5/ True or false ? correct the false . 8m.

Life in Pompeii started 1900 years ago .

There is a shower in the kitchen .

Marten likes same meatballs and pizza .

The astronaut goes round the world times a day ..

.6/ Match the words :. 10m

space	bag
washing	pool
sleeping	station
sun	machine
swimming	glasses

.7/ Write five sentence about your self . what did you do in the week end ? 14m.
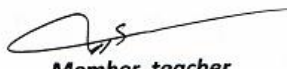
I went to picnic --------

| Q.1 |
| Q.2 |
| Q.3 |
| Q.4 |
| Q.5 |
| Q.6 |
| Q.7 |

Good luck

Teacher in charge

Begard Fateh

Member teacher

Sarfraz Shalal

Appendix H

Q1/ Read the Passage then answer by (Yes or No)   (10 marks)

       Hawler is a fascinating city. Hawler is a capital of Kurdistan. With lots of history and strong Kurdish values. Hawler has some nice places to visit just like garden, shopping mall, swimming pool, and picnic. also the citadel town of Hawler lies in Middle of city center. It's very beautiful and ancient place. The visitors come see this place.

     1. Hawler hasn't any nice place.
     2. My city is fascinating place.
     3. Hawler lies in the east of center.
     4. The visitors come to see this place.
     5. Hawler hasn't garden and shopping mall.

- -

Q2 / Choose the right one :- choose (6)   (12 – marks)

     1. She is .... Her Toys.   (touch , touched , touching)
     2. Sue is watching ....   (film , cartoon , basket ball)
     3. Tom .... run very fast.   (can , could , should)
     4. He .... got any books.   (has , hasn't , haven't)
     5. .... Writing is this?   (where , who's , what)
     6. She eat .... Sandwich.   (a , an , on)
     7. Is it Sue's cap?   (yes, it is , yes, they are , No, it can't)

- -

Q3 / A / Find the Missing letter . choose (5) :   (10 – marks)

     1. wa kman   c othes   app e   choco ate
     2. com   sho s   s at   th ir
     3. ba h   ail   shor   s rong
     4. ca ry   f uit   g ey   siste
     5. range   b x   br wn   p st
     6. p rk   pple   b g   bl ck

- -

Q4 / Pronunciation: find the rhyming words?   (10 – marks)

Shoes , look , near , knee , ring

     1. sing
     2. whose
     3. book
     4. wear
     5. key

106

1

Q5 / Match the sentence with answer: (8 – marks)

List A	List B
1.Is the Parrot flying?	A. Yes, It is.
2.Can she play the guitar?	B. My favorite food is pizza.
3.Who's caps are these?	C. Yes, she can.
4.What's your favorite food?	D. It's Tom's cap.

Q6 / Write composition about (yourself) or your friend? (10 – marks)

١. What's your name?

٢. How old are you?

٣. Where are you from?

٤. What's your favorite food?

٥. What's your favorite sport?

....

....

....

....

....

....

Good Luck

Questions	Marks by number	Marks by written	signature	checker
Q 1				
Q 2				
Q 3				
Q 4				
Q 5				
Q 6				
Total				

Set by :
Tara A. Ismail

Member :
Naqib J. Khorshed

2

94

Appendix I

MAP OF THE BOOK

95

96

Appendix J

Kurdistan Regional Government- Iraq
Council of Ministers
Ministry of Education
General Directorate of Erbil Education
Directorate of Programs

No: 2513

Date: 06- Feb- 2017

To: The Directorate of Central Erbil/ Erbil countryside/ Erbil Environs/ Khabat/ Shaqlawa/ Soran/ Rwanduz/ Choman/ Mergasur/ Makhmur / Koya Educations

Subject/ permission

Permit Madih Ahmed, the MA student at Eastern Mediterranean University who carries his research on Testing, in English language teaching. To collect the data, he needs to analyze the test samples of the seventh grade at Erbil Governorate in Kurdistan Regional Government- Iraq.

Kind Regards

[Legitimate Signature]

Faisal Hassan Taha

General Manager Assistant

A copy to:

-The General directorate Assistant

- Eastern Mediterranean Univesity

- The Directorate of Programs

Email: Gdoe.erbil@moe.gov.krd

Appendix K

Doğu Akdeniz Üniversitesi
"Uluslararası Kariyer İçin"

Eastern Mediterranean University
"For Your International Career"

P.K.: 99628 Gazimağusa, KUZEY KIB
Famagusta, North Cyprus,
via Mersin-10 TURKEY
Tel: (+90) 392 630 1995
Faks/Fax: (+90) 392 630 2919
bayek@emu.edu.tr

Etik Kurulu / *Ethics Committee*

Reference No: ETK00-2017-0152   16.05.2017

RE: Madih Asaad Ahmed (15500427)
Department of English Language Teaching

To Whom It May Concern,

As part of the 2016-2017 Spring Semester, pertaining to Master Thesis questionnaires EMU's Scientific Research and Publication Ethics Committee has granted Mr. Madih Asaad Ahmed (15500427), from the Department English Language Teaching Master Graduate Program, to pursue with his survey entitled *Evaluation of The English Language Test Used by Erbil Governorate Basic Schools on Their Content Validity Seventh Grade*. This decision has been taken by the majority of votes. (Meeting number 2017/42-12)

Regards,

Assoc. Prof. Dr. Şükrü Tüzmen
Director of Ethics Commitee

ŞT/sky.

www.emu.edu.tr