



ISSN: 1817-6798 (Print)

Journal of Tikrit University for Humanities

available online at: <http://www.jtuh.tu.edu.iq>
JTUH
 مجلة جامعة تكريت للعلوم الإنسانية
 Journal of Tikrit University for Humanities

Prof. Nahida Taha Majeed

University of Tikrit College of
Education for Humanities

Jasim Khalaf Jasim

University of Tikrit College of Education
for Humanities* Corresponding author: E-mail: [اميل الباحث: amir.albahath@tu.edu.iq](mailto:amir.albahath@tu.edu.iq)**Keywords:**An Evaluation
English Teacher
Tests
Iraqi Schools**ARTICLE INFO****Article history:**

Received 4 July, 2021

Accepted 17 Aug 2021

Available online 30 Mar 2022

E-mail

journal.of.tikrit.university.of.humanities@tu.edu.iqE-mail: adxxx@tu.edu.iq

An Evaluation of the English Teacher-Made Tests for the Fifth Preparatory Class of Iraqi Schools

A B S T R A C T

Evaluation plays an enormous role in the teaching-learning process. It helps teachers and learners to improve teaching and learning. Evaluation is a continuous process and a periodic exercise. In learning, it contributes to formulation of objectives, designing of learning experiences and assessment of learners' performance. It is very useful to bring improvement in teaching and curriculum. It also provides accountability to the society, parents, and to the education system. A teacher-made test is one of the most valuable instruments in the hands of the teacher to achieve his purpose. It is designed to solve the problem or requirements of the class for which it is prepared and to measure the outcomes and content of local curriculum. It is very much flexible so that, it can be adopted to any procedure and material. It does not require any sophisticated technique for preparation. This study aims at:

- 1- identifying a set of criteria for efficient English teacher-made tests.
- 2- evaluating English teacher-made tests of the fifth preparatory class in the light of the criteria developed. The following procedures are followed in order to achieve these aims:
 1. Designing a checklist which includes the most important criteria of efficient tests, namely: validity, reliability, practicality (scorability) and accuracy.
 - 2- Selecting a sample of eighteen teacher-made tests for the fifth preparatory class.

- 3- Evaluating the efficiency of the selected sample of teacher-made tests in terms of the identified criteria.
- 4- Interviewing a sample of EFL preparatory school teachers for revealing the reasons behind the errors that have been found in the evaluated test papers.

The findings of the study identify a number of defects in teacher-made tests including, invalid questions, questions of low validity, affected reliability, affected scorability and inaccurate questions in punctuation and in grammar. Then the study ends up with some conclusions and recommendations for improving teacher-made tests.

© 2022 JTUH, College of Education for Human Sciences, Tikrit University

DOI: <http://dx.doi.org/10.25130/jtuh.29.3.2.2022.22>

تقويم الأسئلة المدرسية لمادة اللغة الإنكليزية للصف الخامس الإعدادي في المدارس الثانوية العراقية
 ا.د. ناهدة طه مجيد / جامعة تكريت كلية التربية للعلوم الإنسانية
 جاسم خلف جاسم / جامعة تكريت كلية التربية للعلوم الإنسانية
الخلاصة:

تلعب عملية التقويم دوراً رئيسياً في عملية التعليم و التعلم فهي تساعد المعلمين و المتعلمين على تطوير العملية التربوية. التقويم هي عملية مستمرة تساهم في صياغة وتصميم الاهداف والخبرات التعليمية

واختبار أداء المتعلمين و تحسين التدريس والمناهج و كذلك تعطي المسؤولية إلى المجتمع والآباء والنظام التربوي بصورة عامة.

يعتبر الاختبار المدرسي واحداً من أكثر الوسائل المتاحة للمدرس لينجز أهدافه. فهو يُصمم ليحل مشاكل ومتطلبات الصف و يقيس النتائج ومحتوى المناهج. تتميز الاختبارات المدرسية بمرونتها العالية لدرجة إن يمكن الاعتماد عليها في أية اجراءات عملية أو انجاز مهام محددة. ومن مميزات الاختبار المدرسي أنه لا يتطلب أية تقنية معقدة من أجل تحضيره.

تهدف هذه الدراسة الى:

- 1-تحديد مجموعة من المعايير الخاصة بالاختبارات المدرسية لمادة اللغة الإنكليزية.
- 2-تقويم الاختبارات المدرسية لمادة اللغة الإنكليزية للصف الخامس الإعدادي في ضوء المعايير المحددة. تم اتباع الإجراءات التالية لغرض تحقيق تلك الأهداف:
- 1-تصميم قائمة بالمعايير الواجب اتباعها في تصميم أي اختبار: الصدق والثبات و العملية (في القياس) والدقة.

- 2-أختيار عينة تضم ثمانية عشر اختبارا للصف الخامس الإعدادي.
- 3-تقويم فاعلية عينة الأسئلة المدرسية في ضوء المعايير المحددة.
- 4- مقابلة عينة من مدرسي اللغة الإنكليزية من أجل التعرف على أسباب الأخطاء التي وُجدت في عينة الاختبارات المدرسية.

هذا و قد أشارت نتائج الدراسة إلى عدد من السلبيات في نماذج الاختبارات المدرسية حيث تضمنت أسئلة ينقصها الصدق و الثبات وعملية التصحيح أيضا متأثرة بهذه العيوب و احتوت ايضاً على أسئلة غير دقيقة إملائياً و قواعدياً. وقد انتهت الدراسة ببعض الاستنتاجات و التوصيات لتحسين الاختبارات المدرسية.

Section One: Introduction

1.1 Statement of the Problem

English is a dominant language and a required international language of communication, science, information technology, business, and entertainment. It is one of the compulsory subjects in all streams. There must be a process of evaluation through which we can judge the degree of success or failure in achieving the goals set for the educational process in general and the process of teaching English in particular. It should support and reinforce aspects of the instructional progress (Gronlund, 1977:1). Ebel (1972, as cited in Majeed, 2007:240) asserts that "to teach without evaluating or testing, the results of teaching would be foolish".

It is known that the curriculum content reflects the educational objectives which are planned in the virtue of several fields, and the cognitive domain enjoys a greater share from other areas by good feature of several factors including the grade level and nature of material content and the levels of students. Therefore, to measure what students acquire of knowledge and experiences, teachers use achievement tests to carry out these purposes. Achievement tests contribute in improving the learning process. These tests are systematic procedures for determining the amount a student has learned (Dudley-Evans and Maggies, 2005:213).

Evaluation is considered as a technical aspect of instruction and an important indispensable element in the teaching/learning process that gives continuous feedback of data to keep this system in progress. Evaluating of teacher-made tests are considered a major point which attracts the interest and attention of educators and experts of measurement and evaluation. It is a necessary state for learning and a vital element in the process of teaching (Sahab and Abduljadir, 2019:5)

However, probably no scientific study has been carried out in order to evaluate the efficiency of teacher-made tests at Iraqi schools. To the best knowledge of the researcher, most of those tests suffer from defects. Therefore, the current study tries to shed light on the efficiency of teacher-made tests for the fifth preparatory class.

1.2 Value of the Study

Evaluation leads to recycling the process of curriculum development, incorporating changes suggested by the results achieved and changes in educational thought. Examinations occupy a significant importance in the educational process, and in account of the students' fate and their future. All schools can get benefit from examinations in any format, yet because changes in the learner which the education seeks to make cannot view them by naked eye, so there should be a way to show them. Examinations are an extreme tool to achieve this aim, and it is very important to reform them always in order to humor the educational process in its progress. Teacher-made tests are considered a training period before the students enter the exam hall of Ministerial Exams. Fifth preparatory class is an important step in the secondary stage because it precedes the final year, i.e. the sixth preparatory class and it also prepares students for the knowledge and basic skills required to meet most of their needs (Jeffs and Smith, 2005:85).

Hence the value of the study stems from the importance of:

1. identifying the points of strengths and weaknesses of the English teacher-made tests of the fifth preparatory class.
2. investigating and improving the educational process of teaching EFL through investigating the efficiency of the given tests.
3. helping the tests designers through reliance on the levels of questions and criteria set by them.
- 4- paving the way for further studies to evaluate other tests concerning **English for Iraq** series.

1.3 Aims of the Study: This study aims at:

- 1-identifying a set of criteria for the evaluation of English teacher-made tests.
- 2-evaluating English teacher-made tests of the fifth preparatory class in the light of the identified criteria.

1.4 Questions of the Study: The aims of the study are supposed to be achieved through answering the following questions:

- 1-What are the criteria that should be adopted for constructing efficient teacher-made tests?
- 2-To what extent these criteria are adopted in

constructing English teacher-made tests?

1.5 Limits of the

Study: The study is limited to the final English teacher-made written tests for the fifth preparatory class for boys and girls in Al-Sharqat district's schools in Salah Eldeen Governorate of both courses during the academic years 2016, 2017, 2018, 2019 and 2020.

1.6 Operational Definitions of Basic Terms

1.6.1 Evaluation: Evaluation is the process of collecting information systematically for the purpose of indicating the merits as well as demerits of English teacher-made written tests for fifth preparatory class in terms of specific criteria.

1.6.4 Teacher-made tests: They refer to the systematic procedures used by EFL teachers for collecting and interpreting the collected data about their students in order to assess their achievement in what they have been learned.

1.7 Plan of the study: The following procedures are applied in the current study:
1-Identifying a set of criteria for efficient English teacher-made tests.
2-Constructing a checklist that includes the identified criteria.
3-Selecting a sample of English teacher-made tests for fifth preparatory stage.
4-Evaluating the collected sample of English teacher-made test papers according to the identified criteria.
5-Interviewing some of EFL teachers and inquiring them about the essential issues related to the process of constructing efficient teacher-made tests.
6-Collecting the required data and stating some conclusions and recommendations.

Section Two: Theoretical Background and Previous Studies

2.1 Theoretical Background

2.1.1 Procedures of Conducting Evaluation in Education

Richards (2017:287) states many procedures used in conducting evaluation, as follows:

1-Tests: Tests can provide a direct measure of achievement particularly if they are based on students' performance.

2-Interviews: They are done with teachers and students to get their views on any aspect of the course.

3-Questionnaires: They can be used to elicit teachers' and students' comments on a wide range of issues.

2.1.2 Teacher-Made Examinations

Centra (1987:51) believes that tests influence both teacher's and student's performance and that multiple-choice tests in a particular tend not to measure complex cognitive student skills. There are a number of faults in the exams, one being that the questions often asked by faculty members do not represent the wide range of the content in a course. White and Delaney (2018:329-332) specify larger and more general errors which threaten the validity or reliability of entire tests, and minor and specific errors which threaten the validity or reliability of a single item, as follows:

A-Larger and More General Errors which include 1-Failing to Adequately Define the Construct, 2-Non-Linguistic Constructs in Language Testing, 3-Construct-Irrelevant Variance and 4- Muddled Measurement

B-Minor Errors which include 1-Word matching 2-Answer Based on a Single Word 3-Negative Phrasing in a True-False Item 4-Testing the Same Information More Than Once and 5- Testing the Item Rather than the Text

2.1.3 Types of Tests: Tests can be classified as follows:

1-According to the test purpose or programme-level decision:

Brown (1996:9) suggests four types of testing programmes. They are: a- proficiency tests, b- Placement tests, c-Achievement tests and d- Diagnostic tests

2- According to the test timing: These tests are used in identifying the intelligence level and they are either to be limited or unlimited timing tests.

3- According to the test administration: Tests are administered individually or in groups. For individually administered tests, the administrator is an integral part of the testing process. Group tests, on the other hand, are somewhat of a misnomer in that the test does not have to be administered with a group of people.

4- According to the answer type: Different types of tests that formed by the answer type such as oral test where Shohamy et al (2017:182) suggest these qualities of oral performance: unintelligible or no response, partially intelligible, intelligible but labored, or readily intelligible, written test or computer rest.

5-According to the evaluation method: (norm-referenced and criterion-referenced tests) Douglas (2010:67) suggests that the goal of a norm-referenced test is to distinguish among the test takers in terms of the ability being assessed. A criterion-referenced test is therefore designed not to measure a wide range of abilities, but to establish that test takers possess at least a minimum level of ability for some purpose.

6- According to the item format type:

a- Objective tests: Grolund (1977:57) states that this type of test can be used to measure a variety of knowledge outcomes. The most generally useful is the multiple-choice item, but other item types also have a place as true-false items, matching items and short answer items.

b- Essay tests: This kind of tests is used to measure higher level of knowledge in the educational taxonomy (syntheses and evaluation). The most notable characteristic of this type is the freedom of response it provides (ibid: 75).

7- According to the test quality: standardized and non-standardized tests:

A standardized test is the category that has been developed from tryouts and experimentation to ensure that it is reliable and valid (Richards and Schmidt, 2002: 510). A non-standardized test is one that allows for assessment of an individual's abilities or performance, but does not allow for a fair comparison of one student to another (Kadam, 2017: 15).

2.1.4 Characteristics of Good Tests

Gareis and Grant (2015:21) state that when a teacher uses a test, quiz, project, performance, or some other assessment to determine the nature and degree of student learning, the teacher must ensure that the assessment is both valid and reliable. Bachman and Palmer (1996:18) and Rogier (2014:4) propose a model of test usefulness that includes: validity, reliability, authenticity, interactiveness, impact, practicality, usefulness, washback, and transparency as cornerstones in testing that help to ensure that a test is solid.

2.1.4.1 Usefulness: A basis for operationalizing this view of usefulness in the development and use of language tests is provided by the following three principles:

Principle 1 It is the overall usefulness of the test that is to be maximized, rather than the individual qualities that affect usefulness.

Principle 2 The individual test qualities cannot be evaluated independently, but must be evaluated in terms of their combined effect on the overall usefulness of the test.

Principle 3 Test usefulness and the appropriate balance among the different qualities cannot be prescribed in general, but must be determined for each specific testing situation. These principles reflect the belief that, in order to be useful, any given language test must be developed with a specific purpose, a particular group of test takers and a specific language use domain, i.e. situation or context in which the test taker will be using the language outside of the test itself in mind (ibid: 4).

2.1.4.2 Validity: According to Bachman (1990:22) validity is the on-going process of demonstrating that a particular interpretation of test scores is justified, and involves building a logical case in support of a particular interpretation and providing evidence justifying that interpretation. While Gronlund (1977:130) defines it as the appropriateness of the interpretations of test results.

2.1.4.3 Reliability: Ormrod (2014:345) states that the reliability of an assessment instrument or procedure is the extent to which it yields consistent information about the knowledge, skills, or characteristics being assessed. "How consistently does it measure whatever it does measure?" The answer of this question is one of reliability.

2.1.4.4 Authenticity: Bachman and Palmer (1996:24) considers authenticity to be an important test quality because it relates the test task to the domain of generalization to which the teacher wants his score interpretations to generalize. The correspondence between characteristics of target language use (TLU) tasks and those of test tasks is the heart of authenticity.

2.1.4.5 Practicality: This study is concerned primarily with scorability as a part of practicality. Fulcher (2010:46) defines scorability as the ease with which test items or tasks can be scored.

2.1.4.6 Washback: It is defined by Brown and Abeywickrama (2010:354) as the effect of assessment on classroom teaching and learning.

2.1.4.7 Interactiveness: According to Bachman and Palmer (ibid: 27) the interactiveness of a given language test task can thus be characterized in terms of the ways in which the test taker's areas of language knowledge, metacognitive strategies, topical knowledge, and affective schemata are engaged by the test task.

2.1.4.8 Transparency: It makes students part of the testing process by ensuring that they understand what the course objectives are and what will be tested, as well as the format of tests and how they will be used and graded (Rogier, 2014:7).

2.1.5 Techniques Used in Test Construction: According to Al-Juboury (2014:62) any test maker who wants to choose testing techniques, must take to his/her account these considerations: subject matter, age, ability and time.

2.1.5.1 Objective Items or Closed Response: Cunningham (2005:63) states that most educational testing specialists prefer to use objective test items rather than constructed response test items.

1-Multiple-choice item: Gronlund (1977:36) states that a multiple choice item consists of a stem (question or incomplete statement), which represents a problem situation, and several alternatives, which provide possible solutions to the problem.

2-True/false items: According to Al-Juboury (ibid: 66) the true-false item consists of a declarative statement that the testee is asked to tick true (✓) or false (x).

3-Matching items: Gronlund (ibid: 57) states that the matching item is simply a modification of the multiple choice form.

4-Rearranging items: The components of the sentence are scattered and the test taker is asked to rearrange them correctly. Heaton (1990:41) states that rearranging items can take several forms, the first of which to consider is the multiple-choice type.

5-Supplying items: Technique used by test maker where the testee is given a group of words followed by incomplete sentences then he/she is asked to supply these incomplete sentences from the given words (Al-Juboury, 2000: 36)

6-Labelling items: Another form of objective items where the test taker

is required to label certain areas of a diagram or picture (Al-Juboury, ibid: 36).

2.1.5.2 Semi-objective items: They combine characteristics of objective items and constructed response questions. Al-Juboury (2014:78) states several advantages and disadvantage for this type of test items:

Advantages

- 1-Its items are easy and quick to be written.
- 2-It is not difficult to score.
- 3-Guessing the correct answer is limited.
- 4-The relatively high objectivity of marking makes the semi objective test reliable.

Disadvantages

- 1-There are a number of possible right answers involve those that have not been considered by the test maker.
- 2-This technique is not suitable for certain areas like literary appreciation.

Among the test techniques used in semi objective tests are:

1-The transformation: Heaton (1990:46) states that this type is extremely useful for testing ability to produce structures in the target language and helps to provide a balance when included in tests containing multiple choice items. It is the one objective item type which comes closest to measuring some of the skills tested in composition writing.

2-Completion: This technique is a useful means of testing a student's ability to produce acceptable and appropriate forms of language (Heaton, ibid: 47).

3-Cloze test: According to Richards and Schmidt (2002:78) a cloze test is a technique for measuring reading comprehension as well as overall language proficiency. In a cloze test, words are deleted from a reading passage at regular intervals, leaving blanks.

4-Short-answer question: Haladyna and Rodriguez (2013:206) state that short-answer question is scored objectively. Sometimes, a set of short answer items requires some judgment to be exercised to decide if the answer is correct.

2.1.5.3 Subjective Tests: Madsen (1983:9) states that subjective tests, like translation and essay, have the advantage of measuring language skills naturally, almost the way English is used in real life, but many teachers are not able to score such tests quickly and reliably.

A subjective test is a test that is subjectively scored (i.e. scored according to the personal judgement of the marker), such as an essay examination (Richards and Schmidt, 2002:524). Kubiszyn and Borich (2003:55) state that a test is considered subjective when two scorers cannot agree on whether an item is correct or incorrect.

Section	Three:	Procedures
3.1 Population and Samples of the Study		

Population refers to a group of individuals who have the same characteristics. The population of the current study includes all the English teacher-made test papers of the fifth preparatory classes at the schools of Al-Shaqat city during the five academic years 2016-2020 whose total number is fifty-seven papers and this number reflects the actual number of secondary and preparatory schools for boys and girls.

A sample is a subgroup of the target population that the researcher plans to study for generalizing about the target population. The sample of the current study involves the teacher-made test papers of the academic years (2016, 2017, 2018, 2019, and 2020). The collected test papers represent 31% of the total number of the fifth year preparatory class tests. The total number of the questions involved in the selected test papers is 142 and 566 items.

3.2 Achievement of Validity and Reliability of the study

An instrument validity is a reflection of how well it measures what is designed to measure. Face Validity is the best type of validity in case of self-rating (Majeed, N. T., 2008:10).

Thus, the initial form of the checklist has been exposed to a jury of experts who some of them are specialized in Education and others, in Teaching EFL, in order to decide its face validity

3.3 Teachers' Interviews

In order to find out the reasons behind any deficiency found in the investigated teacher-made tests, a checklist has been constructed. It includes seven essential questions raised in front of the EFL teachers who have been interviewed by the researcher. Twenty-three EFL preparatory school teachers are interviewed, i.e. ten are from the right side of Sharqat and thirteen are from the left side of Sharqat. Most of the teachers have been teaching for no less than ten years and have a long experience in teaching **English for Iraq Series** at the various years of secondary schools.

Section Four: Analysis of Data and Discussion of Results

4.1 General Description of the Various Questions and items Included in the Collected Teacher-Made Tests Papers: The total number of the evaluated test papers is eighteen including 142 questions and 566 items. By and large, these questions and items do not consider the criteria and sub-criteria of appropriate tests.

Results indicate that there is a wide inconsistency in the types of the given questions as well as the items included within these questions. The involved teacher-made test papers include fifty-three 'short answer' questions, thirty-five 'free composition' questions, twenty-one 'transformation' questions and only fourteen 'filling blanks' questions, nine 'matching' questions, six 'multiple-choice' questions, and only two questions for each of 'completion' and 'spelling'. Moreover, no 'true-false' and no 'rearrangement' question included.

The involved test papers also show inconsistency in the total number of the items included into the various types of these questions. There are 201, 106, sixty-nine, and sixty items include within the questions of 'short answer', 'transformation', 'filling blanks', and multiple-choice', respectively. There are forty-seven items and forty-five items included within the questions of 'completion' and 'matching' respectively. Moreover, only fourteen items included

within each of 'true-false' and 'rearrangement' questions and only ten items included in the 'spelling' questions.

4.2 Evaluating the Efficiency Teacher-Made Test Papers:

4.2.1 Defects Affect the Validity of the Questions

The First Form: 1-The second question does not have clear instructions, Story Time, where the **English for Iraq Series** does not have this title, rather it is related to the previous series "**Iraq Opportunities**". This point also suits the fourth sub-criterion of validity that reads "A test is valid if it includes tasks that have already been performed as a part of the course procedures."

2-In question three, the first item "Layla's sixteen, (make question tag)" and the second item "He is wearing that silly hat. (question)" affect the tenth sub-criterion of validity because the both are questions.

3-The second item of the third question, "He is wearing that silly hat. (question)", has unclear instruction "question". It is better to be: –Change this sentence into question.

4-The twelfth sub-criterion of validity, is defected in the third question, item number three which reads: "She always early on Mondays. (A-gets up B-get up C-is getting up)" as the word "always" may be considered a clue for the right answer.

5-The fifth question affects the fifth sub-criterion of validity. It is so easy that it involves broad instructions and it does not touch the events of the play "Baghdadi Bath", which also suits the fourth sub-criterion of validity.

6-The first question does not require the students to make conclusions but rather the answers are presented clearly in the passage.

7-The fifth item of the third question, "The machine only (work). You (plug) it in. (zero conditional)", has unclear instruction, which affects the second sub-criterion of validity. It is better to be: (Use if in the zero conditional).

8-The compositions are related to the same idea, "Religious Traditions" though each course has at least three compositions, and that affects the tenth sub-criterion of validity.

The Second Form: 9-The first question is considered easy as it does not include deductive answers which affects the fifth sub-criterion of validity. Students do not have a chance to leave any item and that affects the tenth criterion of validity.

10-For the second question, all the items are taken from two texts. The first item is

related to one text and the other three items are related to another text, though each course includes more than three texts, where the fourth sub-criterion of validity is defected.

11-The fourth item of the second question, which reads: "Mention some advantages of the internet." has unclear instruction as it does not specify how many advantages should be mentioned.

12-The third question, item number three, which reads: "She (already pass) her driving test. (pps)", the instruction is unclear and it is better to be: Re-write the sentence in the present perfect, and that affects the second sub-criterion of validity.

13-The sixth item of the third question, which reads: " –People get fat. They never do sport. (Zero conditional) ", has unclear instruction which affects the second sub-criterion of validity and it is better to be (Join the sentences with If according to the zero conditional)

14-Some main components are neglected in this form, as the vocabulary cannot be represented by these two items of the third question, section B, and these items assess one component at a time.

15-Also the Literature Focus has simply broad questions that do not have the main events of the play " Baghdad Bath ".

The Third Form: 16-The whole unseen passage is so easy which affects the fifth sub-criteria of validity.

17-The unseen passage does not have deductive answers, which increases its easiness.

18-The sixth item of the third question "You are very different your brother, aren't you? "is considered as a clue for the fifth item of the fourth question " Ahmed doesn't wear glasses,? (tags)". In addition, the given instruction is not clear.

19-The second item of the fourth question "What time (have / you) it? (past simple) "may mislead the students to believe that it is a present perfect (have\ you), it is better to be:-What time (you / have) it? (past simple).

20-The sixth item of the fourth question, "Birthday (money, card, computer." , is irrelevant to the syntactical structure rather it goes with vocabulary structure.

21-The test paper lacks the main component of the syllabus "Literature Focus" which affects the content validity.

The Fourth Form has no errors that affect its validity.

The Fifth Form: 22-The whole questions suffer from unclear distribution of

marks which affects the reliability of the test paper

23-The second item of the third question section A “The thieves were in a fight. They dropped the bags. (cause and effect)” is found in the unseen passage which make it easy for the students to answer, this point affects the test validity.

24-The fourth item of section B, question three: “–He did not speak The last thirty minutes. (for, since, from)” the right choice is “for” but the sentence is written in the past simple which may make the students choose the wrong option “from” because “since and for” go with the present perfect.

25-Three questions are related to vocabulary “Q3-C, Q4-A and C” , and that affects the test validity.

The Sixth Form: 26-The second question, the students have the chance to leave almost half of the items as they are required to answer only four questions out of the given seven questions, and that affects the validity of the test.

27-The fourth item, “–Where and when was she born? “, is unclear, who is she?

28-The fifth item “, The accident happened when “, is unclear, which accident?

29-In the third question, the fourth item, “ –What did you buy? “reported “ is unclear, i.e. the students do not know who says and to whom, it is better to be: He/She... asked me/her.....

30-Item number eight of the third question, “ Fatima photo was in the paper. (whose, who, which) “ is considered ambiguous and that affects the test validity. It is better to be: “ Fatima,photo in the paper, is my friend.”

31-The tenth item of the third question, “ They went to London yesterday. (where) “ has unclear instruction, it is better to be: either”-Make a question for this sentence by using “ where “ or Where did they go yesterday? (answer the question, To London)

The Seventh Form: 32-Question six of the Literature Focus is broadly structured and it does not have any events taken from the play.

The Eighth and Ninth Forms have no errors that affect their validity.

The Tenth Form: 33-The second item of validity is affected in the third item of the third question, “ –If I knew the answer, I (tell) you.” does not have any instruction of how to answer which makes the students feel confused, it is better to be: -If I knew the answer, I (tell) you. (Correct the verb in brackets)

34-The tenth item of the third question “you/ ever / clean/ that coat? (have)” has unclear instruction, which affects the second item of validity, it is better is to be: “make a causative sentence using have”

35-The sixth item of the fourth question is irrelevant to the question neither to its type, as the question-type is fill in the blanks from a list and the item is completion, nor to its subject as the question is about vocabulary and the item is related to spelling.

The Eleventh Form: 36-The fifth item of the third question “ Y RU angry 2day? “ is irrelevant to grammar, but it is supposed to be within vocabulary question. This affects the sixth item of validity.

The Twelfth Form 37-The fourth item “What did the doctor say was the business man’s main trouble?” is incomprehensible which affects its validity.

38-The tenth item of the third question “Bamboo is” is irrelevant to grammar, but it is rather related to the question of the book texts. This affects the sixth item of validity.

39-The second item of the second question, “Where and when **were** Widad Al-Orfali born?” , is wrongly structured, it is better to be: **was** instead of **were**.

The Thirteenth and Fourteenth Forms have no errors that affect their validity.

The Fifteenth Form 40-In the eleventh item of the third question, “black, yellow, a beautiful, bird, and. (Re-order the adjectives) “, has a hint for the first word “a”.

It is better to be: (Re-order the words to make a meaningful sentence).

41-All the items of the Literature Focus questions are broad and do not touch the events of the play.

The Sixteenth Form: 42-The second and fourth items of the second question B, “2-I bought (a new pink- a pink new) shirt. 4-I have bought (a new Italian / an Italian new) camera.“, test the same point, i.e. to reorder the adjectives.

43-The fifth question “Literature Focus” is broad and it does not have questions that touch the events of the play.

The Seventeenth Form: 44-The first item of the second question “A”, “ She (have) shorter hair. She (look) nicer. (join: if)“, has unclear instruction, to which condition is it corrected?, it is better to be: “Join with **If** according to the first,

second, third.. conditional”.

The

Eighteenth Form: 45-All the items are taken from only two texts while each course contains more than three texts, and that affects the content validity of the test.

46-The ninth item of the third question, “You read my letters. They are private. (Use: mustn’t) “, is so easy and the student does nothing, and It is better to be: (Use: must)

4.2.2 Defects Affect the Reliability of the Test Papers Questions

The First Form: 1-The third question, section A consists of six items and it is given eight marks. This affects the second sub-criteria of reliability since the given mark cannot be distributed fairly among items.

2-The same is with section B, which contains five mistakes needed to be corrected and four marks are specified for this section .

The Sixth Form: 3-The first question consists of five items and it receives eight marks, which affects the second sub-criteria of reliability since scores cannot be distributed fairly among items.

The Eleventh Form: 4-The third question, the stem of the question is stated on one page and the items are on another page which affects its clarity according to the third sub-criterion of reliability.

The Twelfth Form: 5-The fourth question consists of two sections, four items for each section and the whole question receives four marks, which makes it difficult to distribute the four marks to the eighth items.

The Fourteenth Form: 6-Each of the fourth and the fifth questions are given four marks, and the two questions include five items which makes it difficult to distribute the four marks to these five items equally.

The Other Forms are free from errors that affect the reliability.

4.2.3 Defects Affect the Scorability of the Questions

The First Form: 1-Vocabulary is supposed to be given four marks but it is given eight marks at the expense of writing composition which is supposed to be given eight marks while it is given four marks only, and that affects the first sub-criterion of scorability.

The Second Form: 2-Inaccurate distribution of marks among questions which affects the first sub-criterion of scorability.

3-The question of the book texts is supposed to be given eight marks, but it is given four marks only.

4-The vocabulary is supposed to be given eight marks, but it is given four marks only.

5-The Literature Focus is supposed to be given four marks, but it is given six marks.

6-Writing composition is supposed to be given eight marks, but it is given ten marks.

The Third Form: 7-The inaccurate distribution of marks affects the second sub-criterion of reliability. -The supposed mark of the first question (unseen passage) is eight while it should be scored out of ten.

8-The supposed mark of the second question (book texts) is eight while it receives ten marks.

9-The supposed mark of the third question (vocabulary) is four while it is of five marks.

10-The supposed mark of the fourth question (grammar) is eight while it is given five marks.

11-The supposed mark of the fifth question (composition) is eight while it is of ten marks.

12-The first question, section A, unseen passage, has inaccurate distribution of marks, i.e. the question should be given eight marks but it is given four marks only! Which affects the reliability of the test.

13-Also section B of the first question “book texts“ is of eight marks but it is given four marks only!

14-The second question should be given four marks but it is given six marks, four marks for A and two marks for B.

15-The third question “grammar” is of eight marks but it is given twelve marks at the expense of the other questions.

16-The fourth question should be given four marks but it is given six, two for A and four marks for B.

The Fifth Form: 17.The fifth form consists of eleven questions and none of them has clear distribution of marks at all, which affects their scorability as well as their reliability.

The Sixth Form: 18-The first question is supposed to receive eight marks but it is given five marks, and that makes it difficult for the students to guess the mark specified for each item and impossible for the rater to distribute the eight marks equally to five items.

19-The fourth question is supposed to receive four marks but it receives eight marks.

The Seventh Form: 20-Each question is given eight marks which makes the total marks forty-eight while the total mark of the written test is supposed to be forty.

The Eighth Form: 21-The test paper has inaccurate distribution of marks:

22-Each of the first, second and third questions is scored out of five whereas, they should be scored out of eight.

23-Each of the fourth and fifth questions is scored out of ten whereas, they should be scored out of four only.

24-The sixth question, which should be given eight marks, is given ten marks.

The Ninth Form: The distribution of marks is inaccurate: 25-The first question, section A should be given eight marks but it is given five marks.

26-Section B should be given eight marks but it is given three marks.

27-The third question should be given four marks but it is given ten marks, five marks for A and five marks for B.

The Tenth Form: 28-No marks have been specified to the given questions except the third question which is given ten marks while it is supposed to receive eight marks.

29-The fourth question, of vocabulary, is supposed to receive four marks, the students are asked to answer five items, the difficulty is how to distribute the four marks equally.

The Seventeenth Form: 30-The second question, neither A nor B, has been given any marks.

The Other Forms are free from any errors that may affect their scorability.

4.2.4 Defects Affect the Accuracy of the Questions

The First Form: 1-The fourth item of the second question, which reads: "Mention some advantages of the internet?" ends with a question mark while it is a declarative sentence, and this affects the second sub-criterion of accuracy. Moreover, the first letter of the word 'internet' should be capitalized.

The Second Form: 2-The words "thousands buckets" which appear in the last line of the unseen passage should be written correctly "thousands of buckets".

Moreover, the whole should be ended with a full stop.

3-The second and third questions have more than one item to be included in one line, which affects its accuracy.

4-In the unseen passage the proper noun “ sally “ is written in a small letter instead of “ Sally “ and that affects the second sub-criteria of accuracy.

5-The last line of the unseen passage, which reads: “The man wasn’t hurt. But he swallowed a lot of water” should be written correctly as follows:: ”The man wasn’t hurt, but he swallowed a lot of water.“.

6-The first question is not numbered which affects its accuracy.

The Third Form: 7-There are some errors that affect the test accuracy as in the third line of the unseen passage: “So the government told everyone in China to kill birds. this plan worked well.” There is no need for putting a full stop and even though the full stop must be followed by a capital letter. It is better to be: to kill birds and this plan worked well.

8-Also the phrase “ the farmers produced more food It must be: The farmer

9-The fourth item of the first question “The farmers killed China fifty years ago. (T/F)” cannot be seen as a complete sentence semantically.

10-The fifth item of the first question “The plants were eating by the birds. (T/F)” is intended to be passive but it is not. It should be: The plants were eaten by birds.

11-The first item of the second question “Which animal eating cows?” is constructed wrongly, it is better to be:–Which animal is eating cows?

12-The second item of the second question “Which animal is eating by some people?” is intended to be passive but it is not. It should be: –Which animal is eaten by some people?

13-The sixth item of the second question “There are less than few hundred of which animal?” is not understandable, it is better to be:–How many rote whale are there now? Or – Which animal is endangered?

The Fourth Form: 14-There is a spelling mistake in the verb “phone” which appears in the last line of the unseen passage and it is supposed to be in the past form “ phoned “.

The Fifth Form is free from errors that affect its accuracy.

The Sixth Form: 15-The distribution of marks and some notes are written by hand which makes the test papers look untidy. It is better to supply testees with well

printed test papers.

16-In the unseen passage the word “light” in the second line, should be plural “All the street lights”.

17-In the second item, the word “Al-orfali” is supposed to be started with a capital letter “Al-Orfali” as it is a proper noun.

The Seventh Form: 18-The last line of the unseen passage, the phrase “He wear his black suit” has a syntactical error, it should be “He will wear his black suit “

19-The last word in the unseen passage “bloggs” is supposed to be written as a proper noun “Bloggs”

20-The second item of the sixth question, “From where did the term drama come from? “, it should be: “Where did the term drama come from? “.

The Eighth Form: 21-The fifth item of the fourth question, “You are very, you have good job and lovely family.“, lacks articles in order to be “You have a good job and a lovely family.“

22-In some questions, one line has more than one item as in the first and second questions.

The Ninth Form: 23-The first question, section C, the first item “What are the **elemants** of theoretical performance?”, the word “**elements**” is mistakenly written.

The Tenth and Eleventh Forms are free from any errors that affect their accuracy.

The Twelfth Form: 24-The first question, line number four “ I don’t have any Time for hobbies” where the word “ Time “ starts with a capital letter which affects the second sub-criterion of accuracy.

25-The sixth item of the third question, “I was reading a magazine the phone (ring). (Write the correct form of the verb)”, lacks a conjunction to join the two sentences. It is better to be: “As/While I was reading a magazine, the phone (ring). Or I was reading when/ and the phone (ring).

The Thirteenth Form: 26-The third item of the second question, “What did Mr. Betty Paton see her son Max **did?**”, has a grammatical error which is “ **did** “, the right form is either **do** or **doing**.

27-The fourth item of the second question, “In 2005, four times had the top prize of a million pounds been won in Britain. (T / F)” is structured wrongly. The correct form is “In 2005, the top prize of a million pounds had been won four times in Britain”.

The

Fourteenth Form: 28-The fifth item of the third question “Where did you born? (Make a reported question” is wrongly structured and also it is not clear for the students who asked and to whom. The correct form is: “Where were you born? (Make a reported question) She / He asked me / her “. 29-The seventh item of the third question “stolen, have, the car, been, yesterday” is present perfect and the hint “yesterday” is for the past simple, it is better to be: “The car was stolen yesterday., or The car has been stolen.” 30-The eighth item of the third question, “I know the man. **his** father is a doctor. (join with whose)“, consists of two sentences separated by a full stop, and the second sentence must start with a capital letter. 31-The third item of the fourth question “this soup is ,it needs some salt. “ starts with a small letter. It should be: “This soup is ,it needs some salt.” 32-The fifth item of the fourth question “They **To** take me to the party”, the word **To** is in the middle of the sentence where it is supposed to start with a small letter.

The Fifteenth Form: 33-In the third line of the unseen passage, the word “ **lay** “ is supposed to be “ **laid** “to concord with the sentence. 34-The first item of the fourth question, “silver, gold and tin are all **typ** of“, should start with a capital letter and comma between Silver and gold, and the word types is wrongly spelt. It should be: “Silver, gold and tin are all **types** of”. 35-In the second item of the fourth question, “Everybody wants to be **suhaila**’s friend. **she** is one of the most girls in my class. “, the proper noun **suhaila** and **she** in the second phrase are supposed to start with capital letters. 36-The second item of the fifth question, “ **who was Jawad Al-Assadi?** “, starts with a small letter and the question is supposed to be it the present simple tense, to be: “Who is Jawad Al-Assadi?”. 37-The question of composition, section “B”, “ **write** a letter to your English-speaking friend **tell** him about **atrip** you went on recently. “ starts with a small letter and the words “ atrip “ have to be written “a trip” and the verb **tell** should be **telling**. The correct form of the sentence will be: “Write a letter to your English-speaking friend telling him about a trip you went on recently”.

The Sixteenth Form: 38-In the fourth line on the unseen passage, “There are however,”, there should be a comma before however.
39-The first item of the second question “A”, “While I (have) lunch, somebody stole my bag (correct) “, there is no full stop at the end of the item.

The Seventeenth Form: 40-The third item in the unseen passage, “Why he didn’t catch many fish there?”, is wrongly structured “Why didn’t he catch many fish there?” in the same item the term “ **there** “ is ambiguous.
41-Section A of the sixth question “Write a complaining letter about a place you stay in” has no full stop at the end of the question, which affects the second sub-criterion of accuracy.

The Eighteenth Form: 42-The fifth item of the third question, “I have been writing four hours. She said she..... (Reported speech)”, lacks a main word “for” and that affects the first sub-criteria of the test accuracy. It should be: “I have been writing for four hours.”
43-The seventh item of the third question, “She’s cut her finger/ wash it and put a plaster on it (use: had better) “, has no full stop at the end of the item which affects the second sub-criterion of accuracy. Also its structure looks weak. It should be: “She’s cut her finger. (**Strong advice using:** *wash it and put a plaster on it*”
44-The eleventh item of the third question, “It’s quite **hard** to word with all this noise. (re-write in formal) “, is irrelevant to grammar, rather it is related to vocabulary.

Then “**word**” is spelt wrongly which affects the second sub-criteria of test accuracy. It should be: “**work**”.
45-The second item of the fifth question, “What did Caliban call?“, has inaccurate structure, it is better to be: “What was Caliban called?“ to take the passive form.

4.3 Results of Teachers’ Interview: Twenty-three teachers of English from both sides of Al-Sharqat city have been interviewed in order to answer the questions of the checklist. The intended questions are supposed to show some of the reasons behind the identified errors committed in the various forms of test papers.

The following are the percentages of EFL teacher’s responses to the raised questions, as shown in tables (2) and (3).
1-Just four teachers (17.3%) consider specific criteria and sub-criteria while writing the questions of their test papers?
2-Ten teachers (43.4%) have been visited by a specialist in teaching EFL.

3-Thirteen teachers (56.5%) have designed their test papers according a specific format.

4-Seven teachers (30.4%) have their test papers checked by a specialist, which refers to a weak evaluative role in the teaching and learning process.

5-Thirteen teachers (56.5%) are affected by the increasing number of students which affect the test-type construction. Because the population growth is faster than school construction, the classes have become crowded which affect the whole learning process including test papers construction and scoring.

7-Only four teachers (17.3%) have attended in-service training courses about the appropriate ways of designing English test papers.

Section Five: Conclusions and Recommendations

5.1 Conclusions

The following points are concluded, in terms of the obtained results.

1. The evaluated teacher-made tests have revealed that some questions lack validity because of their ambiguous instructions, inaccurate wording, unclear response and emphasizing on specific language structures and neglecting other structures. The number of items that have low validity is forty-five which represents 7.9% of the total number of the test items.
2. The number of questions that lack reliability is seven which represents the percentage of 4.9% of the total number of the tests items.
3. Sixteen questions which form 11% are not numbered with any score, twenty one questions which form 14.8% receive marks more than the specified ones, ten questions which form 7% receive marks less than specified marks and the number of the questions that are less scorable is thirty-three which form 23% of the total tests.
4. The number of the items that lack accuracy is forty-six which represents 8.12% of the total items of the test papers.

5.2 Recommendations

In the light of the findings of this study, some recommendations are stated as follows:

- 1-The school teachers of EFL should realize well that the construction of any test has specific features and principles. A well-constructed test

- reflects the examiner's competency in his field.
- 2-Teacher-made tests should be constructed in terms of the criteria of validity, reliability, scorability and accuracy.
- 3-Teachers of EFL should be aware of any weak items when constructing their test papers so that the students' performance in these tests reflects the extent to which they have achieved the instructional objectives.
- 4-The Ministry of Education is requested to take account of the importance of the evaluative studies especially those that concern teacher-made tests. Also it is requested to increase the number of educational supervisors of English especially the various directorates of education of Salah Al-Deen Governorate includes only four EFL educational supervisors.
- 5-EFL teachers are advised to avoid emphasis on certain questions and items at the expense of others while designing their test papers and they should be well-qualified to undertake this task.

References

- Al-Juboury, N. (2014) **A Language Teacher's Guide to Assessment**. Baghdad: Al-Yaqdha Printing.
- Al-Juboury, N. (2000) **A Language Teacher's Guide to Assessment**. Baghdad: Al-Yaqdha Printing.
- Bachman, L. F. (1990) **Fundamental Considerations in Language Testing**. Oxford University Press.
- _____ and Palmer, A. S. (1996) **Language Testing in Practice**. UK: Oxford University Press.
- Brown, H. D. and P. Abeywickrama (2010) **Language Assessment: Principles and Classroom Practices**. UK: Pearson Education press.
- Brown, J. D. (1996) **Testing in Language Programs**. USA: Prentice-Hall, Inc.
- Centra, J. A. (1987) "Formative and Summative Evaluation: Parody or Paradox?" **Journal of New Directions for Teaching and Learning**, Vol. 31, 47-55.
- Cunningham, G. K. (2005) **Assessment in the Classroom. (3rd edition.)** USA: Falmer Press.
- Douglas, D. (2010) **Understanding Language Testing**. London: Routledge
- Dudley-Evans, Tony and Maggie Jo St John (2005) **Developments in English for Specific Purposes**. Cambridge: Cambridge University Press.
- Fulcher, G. (2010) **Practical Language Testing**. London: Hodder Education
- Gareis, C. and Grant, L. (2015) **Teacher-Made Assessment**. London: Routledge.
- Gronlund. N. E. (1977) **Constructing Achievement Tests**. USA: Prentice-Hall, Inc.
- Haladyna, T. M. and Rodriguez, M. C. (2013) **Developing and Validating Test Items**. London: Routledge.
- Heaton, J. B. (1990) **Writing English Language Tests**. New York: Longman Inc.

Jeffs, T. and Smith, M.K. (2005) **Informal Education. Conversation, Democracy and Learning**

(3rd edition.), Nottingham: Educational Heretics Press.

Kadam, R. (2017) **Standardized and Non-Standardized Tests**. India: Education

Press. Available at: (http://www.slideshare.net/ReshmaKadam5/seminar-on-standardized-and-nonstandardized-test?from_m_app=android)

Kubiszyn, T. and Borich, G. (2003) **Educational Testing and Measurement. (seventh edition.)**

USA: John Wiley and Sons, Inc.

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

Majeed, N. T. (2007) "Evaluation of the M.A. Final Achievement Tests in "New Trends to Teaching

EFL." **College of Basic Education Researchers Journal**, AL-Mosul University.

_____ "Evaluation of Students' Teaching of English as a Foreign Language at Primary Schools." **Journal of Tikrit University for Humanities**, Vol. (15) No. (11). available online at: <http://www.jtuh.tu.edu.iq>.

Ormrod, J. E. (2014) **Essentials of Educational Psychology**. UK: Pearson Education Limited.

Richards, C. R. (2017) **Curriculum Development in Language Teaching**.

UK: Cambridge University Press.

Richards, J. C. and Schmidt, R. (2002) **Longman Dictionary of Language**

Teaching and Applied Linguistics, UK: Pearson Education.

Rogier, D. (2014) "Assessment Literacy: Building a Base for Better

Teaching and Learning." **English Teaching Forum**, USA: 52 (3) 2-13.

Sahab, Marwan M. and Abduljadir, Yezen, K. (2019) "Evaluation of Headway (Plus) Course book of EFL Undergraduate Iraqi Students." **Journal of Tikrit University for Humanities**, available online at: <http://www.jtuh.tu.edu.iq>.

Shohamy, E.; Or, I. G. and May, S. (2017) **Language Testing and Assessment**. USA: Springer International Publishing.

White, E. and Delaney, T. (2018) **Handbook of Research on Assessment Literacy and Teacher-**

Made Testing in the Language Classroom. USA: IGI Global.