# Training In-Service Teachers in Test Construction Skills (TCS)

## By

### Dr. Karem Abd-ELatif Ahmed

### Assistant Professor, Faculty of Languages and Translation,
### King Khalid University, Saudi Arabia
### Lecturer, Languages Department, Faculty of Management Sciences,
### Sadat Academy for Management Sciences, Egypt

## Abstract

This study aimed at investigating the effect of training in-service teachers of English in language testing on their acquisition and performance in test construction skills. Fifty male and female in-service teachers of English were randomly chosen from the General Diploma, Faculty of Education, Al-Minia University, Egypt. The present study used a semi-experimental method with a pre-test-posttest, control group design. The experimental group was exposed to the two-month training program (administered in the second semester of the academic year 2017-2018) and the pre-and-post data collection procedures. Instruments employed in the present study included a training program in test construction skills made of a students' book and teacher's guide, a written test in the achievement of (TCS) and a written test of performance in (TCS). Analysis of the data by means of a *t*-test for data from the both the pre-tests and posttests indicated significantly better post achievement gains. The researcher concluded that the proposed program had a positive effect on the in-service teachers' acquisition and performance in (TCS). Finally, the study presented educational implications and suggestions for additional research in relation to teacher preparation programs.

**Key words:** Achievement, Skill, Test, Performance Test, Acquisition test and Evaluation.

# تدريب المعلمين أثناء الخدمة على مهارات بناء الاختبارات

## إعداد

## د. كارم عبد اللطيف أحمد محمد

**استاذ مساعد بكلية اللغات والترجمة**

**جامعة الملك خالد – المملكة العربية السعودية**

**الملخص :**

هدفت هذه الدراسة إلى دراسة مدي تأثير تدريب معلمي اللغة الإنجليزية أثناء الخدمة علي مدي قدرتهم علي اكتساب و استخدام مهارات بناء الاختبار؛ حيث تم اختيار عينة عشوائية قوامها خمسون مشاركاً من معلمي اللغة الإنجليزية العاملين أثناء الخدمة من الذكور والإناث من طلبة الدبلوم العام في كلية التربية بجامعة المنيا ، بمصر. استخدم الباحث في الدراسة الحالية المنهج شبه التجريبي ذا التصميم من نوع اختبار قبلي و بعدي ومجموعة ضابطة. تعرضت مجموعة البحث التجريبية إلى البرنامج التدريبي المقترح ومدته شهران في الفصل الدراسي الثاني من العام الأكاديمي 2017 ـ 2018م، وأدوات جمع البيانات القبلية والبعدية. اشتملت أدوات الدراسة الحالية على برنامج تدريبي في مهارات بناء الاختبار مقسم إلى دليل المعلم وكتاب الطالب، واختبار تحريري في اكتساب مهارات بناء الاختبار واختبار آخر لقياس قدرة المعلمين علي الأداء في استخدام مهارات بناء الاختبار. أظهر تحليل البيانات باستخدام اختبار "ت" لكل من الاختبار القبلي و البعدي تحسناً ملحوظاً في أداء المعلمين لصالح الأداء البعدي. وخلص الباحث إلى أن البرنامج المقترح له تأثير إيجابي على اكتساب المعلمين وأدائهم في مهارات بناء الإختبار. وأخيراً، قدمت الدراسة تطبيقات تعليمية مهمة واقتراحات لإجراء بحوث مستقبلية فيما يتعلق ببرامج إعداد المعلمين.

**الكلمات الدالة:** الإختبارات، إختبار الإداء، الإنجاز، المهارة والتقييم.

# Introduction

Tests are the most common assessment tool in the educational systems in the world. Whether students like them or not, this is an actual fact and they have to deal with them. Many critical conclusions that affect students' lives are taken and decided based on certain tests. Bearing in mind these issues, all possible efforts should be made to help test-takers and designers do well in their tests. Hamzeh (2015) in his research stated that, "although test results are used in a wide variety of ways, many test writers, at the public school level, are not skilled in the construction of exams. Consequently, the test items they produce do not test what the items were supposed to test. Thus, the scores attained from such test items may neither be valid nor trustworthy."(pp:30-40)

Testing in school is usually done for the purposes of assessment; to give students grades (from tests in classrooms) or arrange them in relations of their capabilities (in standardized tests). Henry (2011) mentioned the following benefits to testing:

1. The testing result has a great help in taking decisios for future plans.
2. Testing helps in identifying gaps among learners.
3. Testing helps students to learn more for the next stage in their learning
4. process.
5. Testing develops transfer of knowledge to new situations.
6. Testing provides feedback to teachers.
7. Frequent testing inspires students to study and improve. (pp.45-58)

Instructors in their teaching use two processes to help students acquire long-term learning skills; assessment and evaluation. Assessment gives feedback on knowledge, skills, attitudes, and work products for the purpose of evaluating future performance and learning outcomes. Evaluation decides the level of quality of a performance or result and enables decision-making based on the level of quality established. According to Bordon & Owens (2001) assessment and evaluation are used for different reasons, they do have much in common. Both involve specifying criteria to observe in a performance or outcome. Both request the collection of data and other proof by observing the performance or by looking at the result. Both require a performer who collects information about the performance. Both processes conclude

with a report of the findings which include all the similarities and differences. (pp.63-78)

Stressing the importance of good testing in the teaching and learning process, Adodo (2013) emphasized the fact that the kind of evaluation teachers use in the classroom has far reaching consequences not only for students' achievements but also for the nature of the instructional process and feedback assessment reports on those students. (pp.305-10)

Assessment is an essential aspect of teaching. It is the mission that the teachers carry out. It is the method by which teachers collect data about the outcomes of his/her teaching and later use these outcomes for further improvement. According to Ohuche (1998) assessment is the check which teachers carry out to see if his/her students have learnt what they are expected to learn. Hence, teachers have to be trained how to evaluate their students. (p.2)

## The Importance of Testing

There are several benefits for testing, Akinlua (2001) asserted that one important purpose of testing is identifying students' weaknesses with a view to improving future performance. He also stressed the influence of the growth and development of the perceptive, affective and psychomotor skills on the learner hence, the teachers should show high degree of skills in evaluating perceptive and non-cognitive skills of their students. (pp.51-58)

Teachers have a great role in testing according to Weeden (2000) teachers make professional conclusions about learners' performances in all teaching and learning situations, whether consciously or subconsciously. They use these professional decisions and translate them into feedback about the quality of their students' work. This feedback helps the teachers in their assessment. The principal characteristic of assessment for learning is effective feedback on their progress provided by teachers to learners. The value of the feedback is dependent on two factors: first, the quality of the feedback and second, how learners receive and use it. (pp.12-18)

Tests should evaluate what they are designed for, according to Brown (2004) a test is a method of measuring one's ability, knowledge or performance in a given domain. A test is a technique; a tool, a set of methods or procedures, or items which requires performance by the test takers. Some tests are intended to measure general ability, while others measure very specific capabilities or objectives. Testers have to know who the test takers are, what their previous experiences and background are. (pp. 45-69)

Achievement tests are related directly

to classroom lessons, units, or even a total syllabus. Lyman (1998) offered some guidelines when designing achievement testsing which is used most to measure student mastery or knowledge of a certain subject area. First, when planning the test, teachers have to consider examinees' age, ability level, stage of development, culture, etc. These factors will affect construction of learning goals or outcomes, the item formats selected, how items are written, and test length. Second, it is necessary that the content, intellectual or thinking skills psychomotor skills, and attitudes to be assessed are fully and clearly identified and written. This is done through learning outcomes. Third, the test item specifications should be written. Fourth, a table of specifications should be developed which integrates the content, psychomotor skills, and attitudes to be assessed with the intellectual skills and selected test item formats, including the number of items. (pp. 21-26)

## Test Construction Skills (TCS)

Our teachers need to be efficient in test construction, for Silker (2003) being skillful in test construction helps teachers to construct tests with accuracy, suitability of language-use, objectivity. We can not expect teachers to be experts in educational assessment and evaluation to build valid and reliable tests, but there are basic test construction skills that every teacher should possess to construct

tests. These skills enable teachers to structure items, get clear and concise answers from students; and make tests that will be suitable for learners of different ages, abilities, and genders. (pp.204 -226)

Teachers should be skillful in constructing tests because deficiency in test construction skills on the part teachers may result in false assessment of students' outcomes. Paulson ( 2003) thought that the incompetence in test construction resulted from a malpractice in school examinations by teachers in schools. Our teachers should be qualified in constructing tests, Adodo (2001) stated that "teacher should be competent in their subject and in constructing tests so that they can handle materials, equipment competently." (pp.305-319)

Teachers should posses not only the characteristics of good teachers but also the skills of test construction skills, According to Akinlua (2001) the success of any human attempt is closely related to the characteristics of the people who perform these tasks. In any country the ministry of Education is seeking the availability of qualified, committed, devoted, and vocationally spirited teachers to achieve the schools' objectives and national goals on Education. Teachers may be qualified but not efficient. Learning will be ideal when teachers are well trained to be qualified and efficient in their subjects.

Tests have different types, according to Mehrens & Lehmann (1984) there are mainly two kinds of tests. The first one objective tests; matching, true-false, and multiple-choice etc. The second one essay type items. In the second type the students are required to select, arrange, organize, and express ideas or to create original explanations for problems. There is an important issue to consider concerning objective questions which is time. For example, while a multiple-choice exam is graded quickly, the writing of such questions could be time consuming. In contrast, an essay question might be very easy to write, but will require a lot of time to grade. Other factors to consider include the number of students to be tested, the testing place, and the teacher's ability to write diverse types of items.

Certain conditions are required to construct true/false questions efficiently, according to Oosterhof (1994) true/false items encourage predicting because there is an equal chance that either answer will be correct, and a great chance of a correct guess than a multiple_choice item. That means True/false items can be answered correctly without any knowledge of the subject matter being tested because of grammatical clues. To avoid this, Thorndike (1991) suggested some rules in writing true/false items. First, teachers should make sure that the item is clearly true or false. Second, they should avoid the use of exact determiners (an accidental clue to the correct answer). Third, they should avoid the use of negative statements and predominantly double negatives. Fourth, teachers should limit each item to a single idea. Fifth, teachers should make true and false statements nearly equal in length. And finally, teachers should use an equal number of true and false statements.

On designing multiple-choice items, Popham, (2000) mentioned these guidelines:

- Learning outcomes should be reflected in all item construction.
- In the item stem, give a clear stimulus or problem. Each item should address only one central topic.
- The item stem should be written in positive language.
- Make sure that all the items are grammatically right and that answer choices are grammatically parallel.
- Ensure that there is only one right answer per item.
- Avoid using "all of the above," and use "none of the above" unless there is no other choice.
- Make sure that each item stands on its own and is not connected with the following item.
- Avoid any clues that enable students to eliminate incorrect options or choose the correct answer without knowing the content. (pp. 242-250)

On writing matching items Gronlund, (1998) offered some guidelines for istructors to follow:

1. Use only homogeneous content for a group of premises and responses.
2. Make each group short, but make sure that each set is complete. Seven premises and 10-12 choices should be the maximum for each group of matching items.
3. To reduce guessing, list more responses. This prevents students from gaining points through elimination.
4. Put responses in numerical or alphabetical order.
5. Avoid breaking matching items across pages.
6. Use applicable grammatical and syntax rules.
7. Make responses short by using key specific words. (pp. 86-87)

As for Writing Completion and Short Answer Items, Kubiszyn & Borich, (1996) gave the following guidelines for examiners:

- State the statement or question clearly so that there should be only one concise answer of one or two words or phrases.
- Use accurate statement or question wording to prevent ambiguity.
- A complete question is recommended.
- Use one, unbroken blank of enough length for the answer.
- Put the blank at the end of the question or the statement.

- For incomplete sentences, decide a key word or words to be the missing element(s) of the sentence.
- Avoid the following clues e.g., articles such as "a" or "an".(pp. 99-100)

Goals and objectives of designing the tests should be clear in the minds of the teachers, Angelo and Cross (1997) asserted the effectiveness of teachers in building their tests according to them teachers need first to make their goals and objectives clear and then to get specific, complete feedback on the extent to which they are achieving those goals and objectives. This can happen when they are trained to construct effective tests. Effective tests will not only help the teachers but also the students. Students need to improve their learning and receive suitable and focused feedback early and often; they also need to learn how to assess their own learning. (pp. 35-56)

There are some factors should be considered when designing tests, Jacobs (2004) asserted the importance of class size as an important factor in taking the decision about test format. For example, it is very hard to give essay tests when there are (400) students or more in the class because the scoring will be time consuming. A survey of 1100 professors showed that class size is the factor that professors consider to be the most important when deciding the test layout.

Two-thirds of the sample surveyed said they liked the essay format but could not use it because of the size of their classes. They used essay tests only when their classes were small. (pp. 5-12)

## Definition of terms:

To facilitate reading of this study, the following definitions were adopted:

### *A test:*

Mehrens (1989) stated that "a test is any method (written, observational, or oral) utilized to collect data for assessment and evaluation purposes".

Brown (2004) defined a test as "a method of assessing a person's capability, knowledge, or performance in a given area. In other words, a test is a device, a set of procedures, or items that requires performance on the part of the test participants. "This study adopted Brown's definition.

### *Achievement tests:*

To Airasian (1997) "achievement tests are used to measure examinees' present knowledge and skill levels. According to the present study "Achievement tests are summative tests devised to measure how much of a language learners have learned with reference to a specific course of study or programme of instruction. They help the teachers judge the success of their teaching and to recognize the weaknesses of their learners."

### *A performance test:*

Kennison (2013) defined a performance test as "a test in which the ability of applicants to perform particular tasks, usually related with job or study requirements, is measured."

### *An acquisition test:*

Friederici (2011) defined an acquisition test as "the process by which teachers measure the capacity of the learners to perceive and comprehend language, as well as to produce and use words and sentences to communicate as Language acquisition is one of the essential human traits."

### *Evaluation:*

Evaluation is defined by Staff (2011) as "the collection of descriptive and judgmental information necessary to make effective decisions in certain areas of study."

### *A Skill:*

Mojica (2010) stated that "skill provides an individual with the essential knowledge that enables him to operate confidently, effectively and independently in school, life and at work. Individuals who possess these skills are able to participate and progress in education, training, employment"

## Needs for the study:

Reviewing the literature and conducting informal interviews with teachers of English as well as inspectors

of the English language and some staff members helped the researcher reach the need for the present study. Other studies confirmed the need for more research, among them; Silker (2003), Paulson, ( 2003), Eid (1992) and Anwar (2002). All of this helped the researcher to conclude the following:

1- Teachers are in bad need of more training in constructing tests for two reasons:

a- Teachers construct over half of the tests in their classes during the school year according to the inspectors of English.

b- Some of the inspectors of English reported that teacher constructed tests regulary contain item - construction errors.

2- Teachers in this study and many like them were employed as teachers of English, and they had not graduated from a Fculty of Education nor did they receive any formal training concerning language testing.

3- These teachers were only exposed to test construction by their colleagues informally. The researcher chose the present sample for the following reasons:

a- They are actually teachers and employed by the Ministry of Education. They are responsible for teaching and evaluating their classes.

b- These teachers have not graduated from a Faculty of Education so, they have not been prepared to construct tests.

c- Most of their experience in test construction was acquired through their senior teachers or their colleagues. The present study tried to achieve the following aims:

i. Deciding the test construction skills needed by in-service teachers of English.

ii. Enhancing acquisition of these skills by in-service teachers of English.

iii. Enhancing performance of these skills by in-service teachers of English.

## Statement of the problem:

The problem of the study can be specified in the following main question:

How effective would training in-service teachers of English be in the acquisition and performance of test construction skills by these teachers? More specifically, the study tried to answer the following questions:

1- What are the test construction skills needed to be developed by in-service teachers of English?

2- What would the feartures of an effective program of language test construction skills that could be used in training in-service teachers of English be like?

3- How far would a training program in test construction skills enhance the acquisition of these skills by in-service teachers of English?

4- How far would a training program in test construction skills enhance performance of these skills by in-service teachers of English?

## Hypotheses of the study:

The following hypotheses were tested:

1- There would be statistically significant changes between the mean scores of in-service teachers on the pre- post test of the acquisition of test construction skills, favoring the post application.

2- There would be statistically significant differences between the mean scores of in-service teachers of English on the pre- post test of the performance of test construction skills, favoring the post application.

## Delimitation of the study:

The study was delimited to fifty male and female in-service teachers of English. They were randomly selected from a large group enrolled in the General Diploma, Faculty of Education, Al-Minia University, Egypt. The test construction skills involved in the program are limited to the following skills and sub-skills:

A- Grammar tests:
1- Error location.
2- Multiple choice items.
3- Re-arrange items.
B- Vocabulary tests:

1- Synonyms and Antonyms.
2- Derivatives.
3- Filling in.
C- Spelling tests:
1- Roots and suffixes.
2- Missing letters.
D- Comprehension tests:
1- Skimming and scanning questions.
2- True/false items.
3- Ordering given information and details.
E- Writing tests:
1- Writing an outline for a paragraph.
2- Writing a paragraph.
3- Writing an outline for an essay.
4- Writing an essay.

## Material and Method:

The current study started in the first term of the academic year (2016 - 2017) and lasted ten weeks, four hours a week. Two times a week were devoted to teaching and training the experimental group in test construction skills.

## Participants:

Fifty in-service teachers enrolled in the General Diploma in the Faculty of Education in Minia University were randomly selected for the present study. The participants represented most of the towns of Al-Minia Governorate. They had graduated from different colleges. Thirty five had graduated from the Faculty of Arts, English department,

Minia University, and were teaching in preparatory stage. Ten had graduated from the Faculty of Arts, Psychology Department, Minia University and had been employed as teachers of English in the primary stage for seven years. Five had graduated from the Faculty of Commerce, Bani Suef University and had been employed as teachers of English for seven years in the primary stage. During informal interviews, they stated that they were in a great need of training in English language testing skills because they had never received any such training.

## The instructor:

The researcher is the one who taught the program for the experimental group. He also administered the tests.

## Variables of the Research:

- The independent variable is: Training fifty in-service teachers of
- English enrolled in the General Diploma in the Faculty of Education in
- Minia University in test construction skills.
- The dependent variables are:
- Teachers' acquisition of test construction skills.
- Teachers' performance of test construction skills.
- Tools of the research:
- The acquisition of construction skills test.

- The performance skills test.

## Test Construction Skills: (prepared by the researcher).

Reviewing the literature and previous studies helped the researcher in constructing and compiling the test construction skills. The researcher compiled a long list of test construction skills with a grading scale (Important, Yes, No) to be validated by methodology experts as well as language experts and in- service teachers of English. The aim was to collect information about whether these skills are important and required to be taught or not. The researcher distributed the list to a group of (50) in-service teachers of English, (7) staff members in the Faculty of Education, and (5) inspectors of the English language. After analyzing the results, the researcher compiled this final list. They recommended teaching in-service teachers a course to be more fruitful.

Instructional units in test construction skills (prepared by the researcher).

Designing the instructional units followed these procedures:
- Building and compiling the list of test construction skills:
- Reviewing the literature.
- Previous studies.
- Informal interviews with in- service teachers of English enrolled in the General Diploma in the Faculty of Education in Minia University.

- Results of the questionnaire distributed to the (50) in-service teachers, the (7) English professors and (5) English inspectors.
- Deriving and stating the general and behavioral objectives of the instructional units.
- Evaluating the frame of the instructional units by a jury of (5) English

    Professors to decide to what extent objectives, activities and tools of evaluations suit objectives and the subjects of the study.

- Building of the whole instructional units, judging the general form and content.
- The instructional units are divided into two parts:

    Part one: A theoretical background on testing skills.

    Part two: Explaining the skills.

- Designing the program went through the following procedures:
- Preparing the framework of the program: The framework of the program included general and behavioral objectives, content areas, teaching tools, teaching techniques activities and evaluation items.
- Judging of the outline of the program by a jury of English language specialists to decide: how far objectives, content areas, and evaluation   tools suit each other and

how far the program suits the subjects of the study and then suggested modifications were made.
- Building the whole program.
- Judging the whole program by the same jury for its general form and content.
- The content of the program includes 2 parts. It contains 5 lessons and 15 sub-skills. The program begins with general objectives and branches out into number of behavioral objectives. Each lesson starts with a number of behavioral objectives and includes a guided practice and an independent practice. In each lesson the teacher models the skill using the think aloud method. The lessons end with evaluation exercises to measure the in-service teachers'progress. The researcher adopted the material of the program from different resources.
- The program includes two units, each unit dealing with a certain topic as follows:

Unit one: "Theoretical part" "An introduction to Language Testing"

 Lesson one: "Evaluation"

- Definition of evaluation.
- Importance of evaluation for both teachers and students.

 Lesson two: "Testing"

- Test objectives.
- Characteristics of a good test.
- Kinds of tests.

Unit two: Construction of tests according to language skills.

Lesson one: Grammar tests.
    1- Error location.
    2- Multiple choice items.
    3- Re-arrange items.

Lesson two: Vocabulary tests.
    1- Synonyms and Antonyms.
    2- Derivatives.
    3- Filling in.

Lesson three: Spelling tests.
    1- Roots and suffixes.
    2- Missing letters.

Lesson four: Comprehension tests.
    1- Skimming and scanning questins.
    2- True-false items.
    3- Ordering given information and details.

Lesson five: Writing tests.
    1- Writing an outline for a paragraph.
    2- Writing a paragraph.
    3- Writing an outline for an essay.
    4- Writing an essay.

## Evaluation technique:

Two types of evaluation were used in the present research: formative and summative. The procedures of formative evaluation consisted of a set of questions given to the students after each lesson to assess the students' acquisition. The second type of evaluation is summative taking the form of the post administration of the tests.

# Duration of the suggested program:

Two months of four hours a week, two sessions a week. A total of (32) hours.

- Instructional Aids:
- A computer or a laptop for showin data.
- Classroom board.
- The student book.

**Validity of the suggested program:**

To verifying the validity of the program, a copy of the objectives, the program, and the tools teaching techniques were submitted to a jury panel of to determine the face validity of the program and to decide on:

- Deriving and stating the general and behavioral objectives of the program.

- Judging the validity of the objectives by a jury of five staff members as for the following:

- Linguistic stating of the items.

- Appropriateness of behavioral objectives to the general goals.

- How far the objectives can be achieved.

- How far the objectives suit the subjects of the study.

- Appropriateness of the methodology used in the teaching.

## Piloting the program:

The researcher conducted a pilot study that lasted one month. The researcher implemented the first six lessons of the program. The pilot study aimed at ensuring the simplicity of instructions, suitability of the linguistic level of the material to the participants and determining the time the teachers needed to complete each lesson, as well as the approximate time needed for the whole program and the tests. Tools of the study were also administered to the pilot group.

An Achievement Test in the acquisition and performance of test construction skills (prepared by the researcher):

## Objectives of the test:

It was designed to evaluate students' achievement level in the acquisition and performance of test construction skills.

## Construction of the tests:

1-  An achievement acquisition test of language testing (prepared by the researcher). It is one- hundred-item test designed to assess the acquisition of test construction skills by in- service teachers of English language. It is divided into two parts with different guidelines. The first part contains (70) true/-false items while the second part contains (30) completion items. These items represent the specific objectives of the program and measure the acquisition of test construction skills. (knowledge). Guidelines are written in English, they are brief, simple to realize and without any possible ambiguities. According to the opinion of the jury and confirmed by the pilot study, time selected is two hours.

(The tests are attached at the end page 17)

2-  A test of performance in test construction skills (prepared by the researcher)

A selected reading passage was selected titled (Tests). It was used as a base for assessing in-service teachers' ability to construct different   test types taught and trained in. This test measures the ability of the in- service teachers to apply what they learned about test construction skills in actual situations. The subjects were instructed to construct one example of each test skill. The passage readability and content validity were judged by a jury panel. Time allowed was (3) hours. The time was decided by the jury and the results of the pilot study. The total score is one hundred points. Each example has five points. There were three raters to score this question, and the total score was the the sum of the three divided by three to arrive at the average.

 Procedures of designing the test:

1- Identifying the objectives of the t
2- Suggesting items for the test and judging them by jury members.
3- Modifying the test according to the jury members' suggestions.
4- Piloting the test to ensure the clarity of instructions, suitability of linguistic level to the subjects and the suitability of time allowed.

Pre-Testing:( Piloting the tests)

## Validity:

A jury of (5) college staff members examined the tools of the study and approved their face validity and suitability to the level of the subjects and to the set objectives.

## Reliability of the test:

In order to establish its reliability, the tests (Acquistion and performance of test construction skills) were administered to a pilot sample of (30) in-service teachers enrolled in the General Diploma in the Faculty of Education in Minia University. One month later the test was re- administered by using the test retest method. The reliability coefficient of the test is (0.78). This percentage is acceptable.

## Instructions of the tests:

Test instructions are written in English. They are brief, simple to understand and free from any possible ambiguities.

## Scoring the tests:

One mark was allotted for each correct answer of the acquistion test while for the performance test (5) marks were allotted to each skill (15 skills). There were three raters, and the average was taken.

## Time of the tests:

The time of the tests was decided by the jury and according to the results of the pilot study: Two hours for the acquistion test and three hours for the performance test.

## Results and Discussion

### Testing hypothesis one:

Hypothesis one predicted significant differences favoring the post achievement of the teachers of English in the acquisition of test construction skills. The "t" test was used to analyze obtained data. See table (1) for a summary of the results.

## Table (1)

### t – Test Results of the Mean Scores Obtained by in-service Teachers on The Pre- Post Test in The Acquisition of the Test Construction Skills.

| Administration | No. of Subjects | Means | Standard Deviation | "t" Value | Statistical significance |
|---|---|---|---|---|---|
| Pre | 50 | 40.87 | 8.99 | *11.66 | Significant |
| Post | | 66.44 | 11.22 | | |

"t" value is significant at (0.05) level-two tailed test. This means that training in-service teachers of English in test construction skills was effective in enhancing their acquisition of test construction skills. The subjects of study showed a remarkable improvement in the post test. The "t" value is 11.66. Thus, the first hypothesis is confirmed.

## Table (2)

### t – Test Results of the Mean Scores Obtained by in-service Teachers on The Pre- Post Test in the performance of the Test Construction Skills.

| Administration | No. of Subjects | Means | Standard Deviation | "t" Value | Statistical significance |
|---|---|---|---|---|---|
| Pre | 50 | 37.9 | 9.21 | *12.66 | Significant |
| Post | | 65.44 | 11.12 | | |

**Testing hypothesis two:**

Hypothesis two predicted a significant difference between means of scores obtained by the participants in the post measurement test construction skills (performance). Analysis of the results of the data using "t" test showed a significant difference favoring the post achievement of the students. "t" value (*12.66) is significant at level (0.05). This means that instruction and training in-service teachers of English had a good effect on improving their performance in test construction skills. Thus, hypothesis two is affirmed.

## Discussion

The results of the present study indicated that there is a significant difference in the participants' acquisition and performance of test construction skills. Participants acquired and employed significantly more skills after they had received instruction in test construction skills. First, before starting the instruction concerning the theoretical part, the participants showed a lack of knowledge concerning the acquisition of testing skills. That appeared in the pre-tests. The participants' answers showed that they had no idea about the conditions of a good test, a diagnostic test, an achievement test, a performance test or the best way to test the different skills. Some of them heard about these kinds of skills and sub-skills for the first

time. Their knowledge was limited to the names of some skills; such as multiple-choice questions, true and false questions, and matching questions.

Most of the participants did not know how to construct these types of questions in the right way. They did not know the difference between objective and subjective tests, how to write good essay questions or when to use objective tests and essay questions. After receiving instruction their performance improved significantely. That appeared clearly in the results of their post acquistions tests.

As for the performance of testing skills and sub-skills, to prove the second hypothesis, in the pre-testing the participants could not design the skills or sub-skills in the required way. For skill one; the grammar skill, the participants were asked to construct questions using the sub-skills; error location, multiple choice items and re-arrange items. The participants showed that they had no idea how to design questions concerning the first sub-skill; error location. They were used to writing re-arrange questions and multiple choice questions and even when they were designed wrongly, the instructions were not clear and designed them wrongly. For example, in the multiple-choice questions they sometimes wrote the correct answer longer or shorter than the other options. They used the phrases "all the above" and "none of the above" options. The

problem with these phrases as an option is that they make the item too easy. If the students could recognize at least one wrong option, they could reject it. So, it was considered as a clear clue to the correct answer. In the second skill; vocabulary tests and its sub-skills; testing synonyms and antonyms, testing derivatives and testing filling in. The participants were used to writing questions like fill in the gaps from the given list. Even their questions were full of grammatical mistakes and punctuation mistakes. They could not design questions testing synonyms, antonyms, or derivatives.

These skills were new for them. For the third skill; spelling tests and its sub-skills; roots and suffixes and missing letters, the participants were not able to construct questions testing roots and suffixes. They did not know the meaning of affixes, prefixes and suffixes. Concerning the fourth skill; testing comprehension and its sub-skills; skimming and scanning questions, true-false items and ordering given information and details. The results proved that the paticipants could not design questions testing skimming or scanning. They used to test their students' comprehension using true false questions and Wh. questions. As most of the participants were not specialists in English, these types of questions were new for them. Concerning the fifth skill; writing tests and its sub-skills; writing an

outline for a paragraph, writing a paragraph, writing an outline for an essay and writing an essay, the researcher found most of the participants had no idea about the difference between paragraph writing and essay writing. It was the first time for them to hear about topic sentences and thesis statement. They didnot not even know how to construct such questions.

After the treatment, subjects' performance improved significantely. That was evidenced clearly in their post-performance tests. They could recognize the meaning of a topic sentence, supporting details, thesis statement, coherence and cohesion. In addition, they could recognize the different types of paragraphs and essays such as descriptive, narrative, comparison and contrast paragraphs and essays.

Here are samples of their questions; For example; before the treatment, when the participants were asked to construct questions measuring the grammar of their students, they committed serious mistakes. Here is one of the questions before the experiment:

"Choose between brackets." As we can see the question is too short. The teacher should have stated: "Choose the correct word or words from the brackets." Another example: " Match between the two columns." First, there should be no preposition ("between") here. Second, this question is suitable for

testing vocabulary not grammar. When testing their students' comprehension, the participants committed many grammatical and spelling mistakes at the same time the instructions were not clear. For example:

1- *Answer the following question by reading them and then writing either a simple or compound sentences for each according to the questions?".* "Bear in mind that this is a comprehension question". It is full of grammatical mistakes and it does not test comprehension."

2- Read the following short passage and then decide: What is the best suitable choice from among the given choices; can be the main idea for this passage you have read?

It is a very long question that will make the students forget the objective of the question. The final results of the present study supported some previous studies on test construction. For example, Mordi (2013) indicated that test training can be effective for language teachers. He also asserted the importance of training teachers in how to assess their students and give periodical tests during teaching to observe their students' learning improvement and to provide on-going feedbacks to both teachers and students. Such results also agree with previous findings of Anwar (2002), Ozmen (2011) and Eid (1992) who found that teachers who received training in

test construction skills improved significantly over those who had received no instruction.

# Discussion

It is a known fact that the success of any learning environment is determined by many factors. In my opinion the most important factor is assessment. There should be steps toward enhancing assessment in our schools. The evaluation of pupils' progress is a major aspect of the teacher's job. This study is intended to strengthen the ability of in-service teachers in designing the testing skills as it has a great role in improving teaching/ learning situations for both teachers and students and improving their performance and outcomes. It also important when taking decisions concerning the educational process for both. In-service teachers are in a bad need for continuous training on testing skills. This study proved that training in-service teachers on testing construction skills improved their acquisition and performance of these skills.

## Recommendations of the research

In light of the results and conclusions of the present study, the following recommendations are suggested:

1- In-service teachers of English should be trained in effective test construction skills through a comprehensive program.

2- The objectives of English preparation programs in colleges should be re-examined to include courses in developing test strategies and skills.

3- Regular workshops and seminars should be held to build the capabilities of teachers in using new methods of test constructing and lesson evaluation.

4- In-service teachers of English should be encouraged and sponsored to attend conferences related to testing.

## Suggestions for further research:

The following topics are suggested as areas that need further investigations:

1- This study may be repeated on a big sample or even pre-service teachers of English.

2- The present study focused on test construction skills. Other studies can focus on test use and, scoring, and analyzing test scores.

3- Other studies may deal with testing high level thinking skills.

4- Other studies can be conducted on listening and speaking tests as they are important and need special preparation.

5- Empirical studies are needed to determine the long-term effect of training in effective test construction skills.

6- Empirical research is needed to examine the relationship between testing skills and language proficiency.

## References

Adodo, S. (2013). Correlate of Pre-Service Teachers and In-Service Teachers Perceived and Prioritized Students' Psychological Profiles for The Teaching and Evaluating Basic Science and Technology (BST). *Journal of Emerging Trends in Educational Research and Policy Studies, 4*(2), 305-319.

Airasian, P. (1997). *Classroom Assessment* (3rd ed.). New York City, NY: McGraw-Hill.

Akinlua, J. (2001). Effectiveness of Two Types of Teacher's Remediation on student Achievement in Integrated Science. *Journal of Educational Research and Evaluation, 4*(2), 51-58.

Angelo, T. A. (1995). Classroom Assessment Techniques: A Handbook for College Teachers. *The Journal of Higher Education, 66*(1), 108. doi:10.2307/2943957

Anwar, Z. (2002). The Effect of Training Pre-Service Teachers of English in Language Testing on Their Acquisition and Performance in Test Construction Skills. *Journal of Research and Education and Psychology, Minia University Faculty of Education, 16*(1), 30-38.

Borden, V., & Owens, J. (2001). Measuring Quality: Choosing Among Surveys and Other Assessments of College Quality.

Bott, A. (1996). Testing and Assessment in the Occupational and Technical Education. Boston: Allyn & Bacon.

Henry, M. (2011). Ten Benefits in Tests and Their Applications to Educational Practice" *Psychology of Learning and*

*Motivation*", 55. (5).45-58.

Jacobs, L. C. (2004). Indiana University. Available at http://www.indiana.edu/~best/write_better_tests.shtml Pp. 5-12

Kennison, M. (2013). Introduction to language development. Los Angeles: SAGE Publications.

Kubiszyn, T. & Borich, G. (1996). Educational testing and measurement. New York, NY: Harper Collins College Publishers.

Lyman, H. B. (1998). Test scores and what they mean (6th ed.). Needham Heights, MA: Allyn & Bacon.

Mehrens, W. A., & Kaminski, J. (1989). Methods for Improving Standardized Test Scores: Fruitful, Fruitless, or Fraudulent? *Educational Measurement: Issues and Practice, 8*(1), 14-22. doi:10.1111/j.1745-3992.1989.tb00304.x

Mojica, L. (2010). An investigation on Self-reported writing problems and actual writing difficulties of EFL learners in the beginner's level. *TESOL Journal*, 2, 24-38

Ohuche, R. O. (1998). Continuous assessment for every learner. Onitsha: African—FEP

Mordi, C. E. (2013). Understanding students, Academic Achievement "30th in the Series of Inaugural Lectures of the Delta State University, Abraka, Nigeria.

Oosterhof, A. (2001). *Classroom Applications of Educational measurement*. Upper Saddle River, NJ: Merrill.

Özmen, K. (2011). Washback Effects of the Inter-University Foreign language examination on foreign language competences of candidate academics.

Paulson, T. (2003). The study of teaching. New York: Holt, Rinehart and Winston Inc.

Popham, W. J. (2000). Modern educational measurement (3rd ed.). Needham Heights, MA: Allyn & Bacon.

Silker, R. (2003). Teachers and Tests. London: BasilBlackwell.

Staff, T. (2011). Lessons in EffectiveTeaching. Learning Technologies at Virginia Tech.

Thorndike, L. (1991). Measurement and Evaluation in the Psychology and Education (5th ed.). New York: Macmillan Publishing Company.

Weeden, P. (2000). The Learn Project: Guidance for our Schools on Assessment for Learning. (Available on the QCA website – assessment for learning).