

Evaluation of the M.A. Final Achievement Tests in "New Trends to Teaching EFL"

**Asst. Prof.
Dr. Nahida Taha Majeed**

The 1st Scientific Annual Conference for College of Basic Education (23-24/May/2007)

Abstract:

This study aims at:

1. Setting up appropriate criteria for evaluating the efficiency of the M.A. final tests of "New Trends to Teaching EFL".
2. Evaluating the identified tests on the basis of the developed criteria from the points of view of specialists in Education, Methodology of Teaching English, and Curriculum Construction.

This study confines itself to evaluating the efficiency of end-of-course achievement tests of "New Trends to Teaching EFL" to which M.A. students are subjected at Baghdad and Al-Mustansiriya Universities, during the last seven years: 2000, 2001, 2002, 2003, 2004 and 2005.

The sample of this study implies the written achievement tests of "New Trends to Teaching EFL" subject which are given to M.A. students at the end of the course examinations (first attempts) during the seven consecutive years from 1999-2000 up to 2004-2005. This sample consists of twenty-two test- papers obtained from Baghdad University as well as Al-Mustansiriya University. The test papers include 112 questions.

A specific checklist is constructed for the purpose of evaluating the selected sample of test papers. The checklist includes the following major criteria:

The obtained results indicate that M.A. achievement tests suffer from a clear shortage in their technical aspects.

At the end, the study has stated some points of conclusions and recommendations.

تقييم الاختبارات التحصيلية لطلبة الماجستير لمادة (اتجاهات حديثه في تدريس اللغة الإنكليزية)

أ.م.د. ناهده طه مجيد

ملخص البحث :

تهدف هذه الدراسة إلى:

1. وضع المعايير المناسبة لتقويم كفاية الامتحانات النهائية لطلبة الماجستير في مادة "الاتجاهات الحديثة في تدريس اللغة الانكليزية" بوصفها لغة أجنبية.
 2. تقويم الاختبارات بضوء المعايير الموضوعية و ذلك من وجهة نظر المختصين في التربية و بناء المناهج و طرائق تدريس اللغة الانكليزية.
- تحدد هذه الدراسة بتقويم كفاية الاختبارات التحصيلية النهائية لمادة " الاتجاهات الحديثة في تدريس اللغة الانكليزية" لطلبة الماجستير في تدريس اللغة الانكليزية في جامعة بغداد و الجامعة المستنصرية وذلك للسنوات الست الماضية 2000, 2001 , 2002, 2003, 2004 و 2005.

واشتملت عينة هذه الدراسة على 22 نسخة من الامتحانات التحريرية التحصيلية و المتضمنة 112 سؤالاً قدمت لطلبة الماجستير كاختبارات نهائية للسنوات الست الماضية منذ 1999-2000 ولغاية 2004-2005 في جامعة بغداد و الجامعة المستنصرية.

أما أداة الدراسة فتتمثل ببناء قائمة تحقق "checklist" لغرض تقويم الاختبارات المتضمنة في عينة الدراسة و تتضمن هذه القائمة المعايير الآتية: الأهداف المعرفية التربوية لتصنيف "بلوم" و التي تقسم إلى ستة مستويات هي: التذكر , الفهم , التطبيق , التحليل , التركيب و التقويم ، الصدق ، الشمولية ، الموضوعية ، دقة ووضوح الصياغة ، سهولة القياس ، الإخراج و الطباعة و الشكل.

بعدها تم التأكد من مدى ملائمة المعايير الموضوعية في تقويم الاختبارات التحصيلية لطلبة الماجستير، وذلك بعرضها على عينة من الخبراء و المختصين.

أخيراً توصلت الدراسة إلى عدد من الاستنتاجات والتوصيات التي يمكن أن تفيد التدريسيين في الجامعات العراقية وكذلك المختصين و القائمين على عملية بناء الاختبارات التحصيلية لدى طلبة الدراسات العليا.

Introduction:

Evaluation is one of the components of educational process. It is a necessary condition for learning and an essential element in the process of teaching. In fact, evaluation and learning processes are closely interrelated with each other. When evaluation is inaccurate or absent, the effects upon the entire instructional system and upon learning outcomes are disastrous. According to Ebel (1972: 41) "to teach without evaluating or testing, the results of teaching would be foolish".

The purpose of evaluation is to reflect critically on the effectiveness of personal and professional practice. It is to contribute to the development of good practice (Everitt et al, 1996: 129). Smith (2006) adds that evaluation is not primarily about the counting and measuring of things. It entails valuing- and to do this, specialists have to ensure that this process of looking, thinking, and acting is participative.

Tests are one of the means of evaluation. Evaluation relies heavily on paper and pencil tests, most of which are classroom rather than standardized tests. No other means that are efficient and beneficial have been discovered yet instead of testing. They are directly related to all the factors in the teaching-learning system. It is through tests, the comprehensive assessment of learning capabilities is carried out. They can be used as a part of the instructional strategy, both as a learning vehicle and as a means of reinforcement. But the primary function of tests is to measure the achievement of instructional objectives of the teaching-learning process. Lyman (1971: 4) believes that tests are often used as tools in getting decisions made on behalf of an institution.

Therefore, analysing and evaluating a sample of the final achievement tests at M.A. stage as a reflection of how well these tests measure what they are designed to measure, form the main concern of this study which includes five sections.

Section One

1.1 Definition of the Problem

Higher study constitutes the most crucial stage in the ladder of education and instruction in Iraq. At this stage, graduate students are trained to teach at universities, colleges, or other higher institutions after finishing their M.A. or Ph.D. studies. The worth of the whole educational process contributes to the success of higher instruction in preparing well-trained and specialized college instructors.

With the need for more efficient M.A. students at the various instructional subjects, goes the need for adopting efficient means of measuring their performance and effort in the given courses. Those students are occasionally subjected to quizzes, asked to write scientific essays through consulting recent issues and references related to the intended syllabus items, and prepare weekly and termly papers that can, in one way or another widen their understanding of the instructional subject. All these measurement techniques go hand in hand with the presented instructional materials.

However, assessing the academic achievement of M.A. students at the educational colleges in Iraq, relies heavily on paper and pencil tests. It is through such written tests the comprehensive assessment of 70% of students' learning capabilities is carried out at the end of each course of study. The other 30% of the final students' score is allotted to the above-mentioned assessment techniques which are accomplished during the course of study.

A number of researches have been carried out to investigate the problem related to the analysis and evaluation of classroom tests at the various stages of study, i.e. primary, secondary and college. Nevertheless, analysis and evaluation of achievement tests at M.A. stage so far, has not been subjected to a scientific investigation. As far as the researcher

knows, no previous scientific work has undertaken the responsibility of subjecting any M.A. tests to evaluative study. Hence, the efficiency of such tests is questionable.

This study is an attempt to investigate the efficiency of a sample of final achievement tests for one of the crucial subjects studied by M.A. students who are going to have specialization in Methodology of Teaching EFL. This subject is entitled, "New Trends to Teaching English". Thus, the current study is a pioneer piece of work undertaken to evaluate and consequently increase the effectiveness of such tests.

1.2 Value of the Study

Evaluation facilitates learning and teaching as well as produces records appropriate to the purposes for which records are essential. It also provides continuing feedback into the larger questions of curriculum development and educational policy (Wilhems, 1967: 15). Evaluation as a part and parcel of education is viewed as an integral part of the development or change process (Rubin, 1995: 17-23 and Smith, 2006: 40). It leads to recycling the process of curriculum development, incorporating changes suggested by the results achieved and changes in educational thought (Jeffs and Smith, 2005: 85-92).

The final achievement tests posed to M.A. students represent the most convenient instrument to evaluate the progress achieved by testees in terms of their acquisition of the desired behavioural patterns specified by the various items of the instructional syllabuses. Moreover, unless achievement tests are well constructed and administrated, they may be a source of danger to the whole instructional process (Stiggins, 2003 and Scott, 2005).

In spite of the essence and sensitivity of M.A. final tests, they have never become an area of investigation. More or less, the value of this study can be manifested in the following points:

- 1.2.1** The study is expected to make a number of contributions to our understanding of how far our college faculties have been successful in constructing appropriate achievement tests.
- 1.2.2** The practical value of this study lies in its determination of the efficiency of M.A. achievement tests from the point of view of specialists in the fieldwork of this study.
- 1.2.3** Developing certain criteria for judging the efficiency of the intended tests which will serve as guides for college instructors in constructing their tests.
- 1.2.4** The significance of this study lies in the fact that it tackles testing students' achievement in "New Trends to Teaching EFL" which forms the basic subject studied by M.A. students specialized in Methodology of Teaching English.
- 1.2.5** Since no similar study has undertaken such a task, this study is considered a pilot project undertaken to identify the consistency of final tests with the objectives of M.A. programme.
- 1.2.6** This study will fill a definite gap of information in research and pave the way for further investigation of higher stages (M.A. and Ph.D.) tests.

1.3 Aims of the Study

This study aims at:

- 1.3.1** Setting up appropriate criteria for evaluating the efficiency of the M.A. final tests in "New Trends to Teaching EFL".
- 1.3.2** Evaluating the identified tests on the basis of the developed criteria from the points of view of specialists in education, methodology of teaching English and curriculum construction.

1.4 Limits of Study

This study confines itself to evaluating the efficiency of end-of-course achievement tests in "New Trends to Teaching EFL" to which M.A. students are subjected at Baghdad and Al-Mustansiriya Universities, during the last six years: 2000, 2001, 2002, 2003, 2004 and 2005.

1.5 Definitions of Basic Terms

As used in this study, the following basic terms need to be defined operationally for the purpose of clarity and coherence:

- 1.5.1 Achievement Test:** it is a type of measurement instrument constructed by a college faculty that includes a group of questions or tasks in any given course of study to which M.A. students should respond at the end of that course.
- 1.5.2 Evaluation:** It means the process of judging tests value on the basis of certain criteria specified for the purpose of making decisions about the efficiency of the end-of-course achievement tests.
- 1.5.3 M.A. Stage:** It refers to the educational stage that comprises two years of consecutive study after the B.A. stage. It grants post graduates, specialization in "Methodology of Teaching English".
- 1.5.4 "New Trends to Teaching English":** It is one to the basic subjects studied by M.A. students who are going to have specialization in "Methodology of Teaching English". This course of study acquaints the students with the new theories and trends to teaching EFL, their principles, characteristics, methods and classroom techniques. It also supplies them with pedagogical applications for a variety of these trends in classroom situations.

Section Two

Previous Studies

II.1 Introductory Note

To the best knowledge of the researcher, no pervious study so far, has subjected M.A. or Ph.D. tests to a scientific piece of work. However, there are a lot of studies that have evaluated achievement tests at primary and secondary stages.

Moreover, two previous studies have been accomplished at college stage hence, these previous studies could be classified into two groups, as follows:

II.2 First Group

This group includes the studies which have evaluated achievement tests of different subjects at primary, intermediate, and preparatory schools. As they are not related closely to the present study, they will be stated here without discussion as shown below in Table (1):

Table (1) : Classification of the First Group Studies

subject	Title of the study	Researcher's Name and Years of the Study	No.
Science and Health Education	Evaluating the Written Questions of the Science and Health Education for Fifth Primary School	Al- Aumer, Nasreen (1971)	1
History	History Examination in Group Schools in Lankshiyer	Directorate of Testing in Lankshiyer (1972)	2
English	Evaluating the Annual English Classroom Tests in Grade V of the Preparatory School	Fernandes, H.J. and Issa. (1972)	3
Arabic	Evaluating Written Tests of Arabic Handwriting for Fourth Primary School	Bahri, Muna Y.(1973)	4
English	Evaluation of English Classroom Tests in Grades 4 and 5 of the Preparatory School	Razzaq, F. and Helen Al-Hasan (1973)	5
All School Subjects	A Depth Study of the Ministerial Examinations for the Sixth Primary Schools.	Al-Nasser, B. and Y. Yaqoub (1974)	6
Arabic	Evaluating the Final Tests of Arabic Grammar for Second Intermediate School	Dama'a, Majeed I. (1974)	7
History	A Descriptive Analysis of the Content Validity of the History Examinations of Ethiopian Secondary School.	Abid, Yousif Omer (1975)	8
Arabic	Evaluating Classroom Written Tests in Arabic Grammar for Fourth Primary School in Iraq.	Bahir, Muna Y. and Majeed I. Dama'a (1975)	9
English	The Influence of the In-Service Teacher Training on English Teacher-Made Tests.	Al-Kubaisy, Amir B. (1979)	10

subject	Title of the study	Researcher's Name and Years of the Study	No.
Education	Quantity Analysis of the Final Written Tests for Sixth Primary Level in National and Social Education.	Al-A'ani, Nasar (1980)	11
Geography	Quantity Analysis of the Final Written Tests for Sixth Level in Geography.	Al-A'ani, Nasar (1980)	12
Science	Evaluating the Questions Included in Science Textbooks in the Three General Level of Instruction.	Al-Damardash, Ibrahim S. (1980)	13
Arabic	Evaluating the Final Tests of Arabic Grammar for Fourth Preparatory School in Iraq.	Bahry, Muna Y. (1980)	14
English	The Correspondence Between Teacher's Questions and Student's Answer in Classroom Discourse.	Stephen, R. M. and David C. B.	15
Social Subjects	An Analysis of Questions in the Sixth Grade Social Students Textbooks.	Frankline, Doris M. (1981)	16
Social Subjects	Evaluating Primary School Examination of Social Subjects.	Directorate of Evaluation and Educational Directors/ Ministry of Education (1983)	17
English	The Effectiveness of English General Baccalaureate Examination in Testing Oral Competence.	Al-Nassiri, Nahida T. (1984)	18
Geography	Question Analysis of First and Second Books in Geography in the Light of Curriculum Objects and Learning Conducts.	Abu-Al Ala (1985)	19
All School Subjects	Studying and Evaluating of Public General Examinations of Primary School of all Iraqi Governorates.	General Directorate of Evaluation and Testing (1986)	20
History	An Evaluation of the Final Examination of History for the 1 st and 2 nd Intermediate Class.	Muhammed, H. S. (1987)	21
Chemistry, Physics, Science, Health and Mathematics	Studies in the Quality Analysis of Ministry Examination Papers.	Al-Anee, Nasar et al (1988)	22
Geography	Evaluation of Final Examination Papers in Geography for the 1 st and 2 nd Intermediate School.	Al-Bawi, Abbas (1989)	23
National Education	Evaluation of Final Examination Papers in National Education for the 1 st and 2 nd Intermediate School.	Al-Eesaw, Karim N. A. (1989)	24
Chemistry	Criteria of Good Tests in Analysis Study of the General Examination Papers in Chemistry for the 6 th Class of the Scientific Branch.	Al-Zuhari, Kamil K. (1990)	25
Science	Evaluation of Science Examination Papers of the Final Stage of Learning.	Salam, Safea (1990)	26
Science	Evaluation of Science Examination Papers of the 5 th Scientific Class in Science.	Al-Afoon, Nadea (1991)	27
History	Evaluation and Developing the Ministerial History Examination Papers for the Sixth Grade in Iraqi Preparatory Schools.	Izzat, Sua'd (1993)	28
English	Evaluation of English Teacher Made Tests for the Fourth Year in Iraqi Secondary Schools.	Al-Samarai, Muaid N. (1996)	29
English	A Comparative Study of the Iraqi Baccalaureate and Jordanian General Secondary School Examinations of English.	Mohammed, Emade Al-Din A. (1996)	30

subject	Title of the study	Researcher's Name and Years of the Study	No.
History	Evaluating General History Examination Papers in General Secondary Stage in Yamen.	Al-Salahi, Abdul Salam (1998)	31
English	Evaluation of General English Examination Papers for the Sixth Primary Schools.	Al-Rubai'e, Ayah M. (1999)	32
English	Evaluation of Teacher Made and Central Tests in EFL at the Second Intermediate Stage.	Al-Azizi, Eman M. (2000)	33
English	Validity and Reliability of General Baccalaureate Examinations in Testing the Functional Aspect of English.	Al-Mashhadani, Abdullah M. (2002)	34
English	Evaluating the General Secondary Examination of English (GSEE) of the Academic Stream in Jordan.	Al-Ro'oud, Atallah, A. (2003)	35

II.3 Second Group

The previous studies included in this group have evaluated achievement tests at college stage. These studies will be discussed here since they are, in one way or another, related to the present work as follows:

II.3.1 Jameel, 1982

This study aims at studying the status of the examinations in the area of Geometric instruction in order to identify the obstacles that lead to the decrease of students' scientific level.

The sample of the study includes thirty-five college instructors and forty-five students. A variety of the instruments below are used:

- a- interviewing a number of specialists in the fields of geometric instruction, examination, and evaluation.
- b- analyzing the results reached at by the examinations of the past twelve years.
- c- designing a questionnaire and asking the points of view of college instructors concerning the ways and designs of geometric examinations.

- d- designing another questionnaire in order to ask the points of view of the students of Technology University concerning evaluation and examination.

Results of the study have revealed the following:

- 1- 79% of the college instructors prepare a sample of tests in advance then they choose the appropriate tests to their students concerning time and comprehensiveness.
- 2- 86% of college instructors have a bank of tests for using them in future.
- 3- These instructors prefer to include 39% of their tests, simple, main and sub-questions, 34% of the evaluated tests are comprehensive and 27% are moderate in their level of difficulty.
- 4- 95% of the students believe that objective tests are the best tools of evaluation.
- 5- College students face thirteen problems throughout their examination process.

II.3.2 Faisal, 2002

This study aims at evaluating English achievement tests in grammatical structures of the first year at the College of Arts and the College of Languages in Baghdad University. The process of evaluation is accomplished according to seven criteria developed by the researcher as well as the objectives of the intended course (grammatical structures).

The adopted criteria include: layout, accuracy, Validity, objectivity, reliability, practicality, and sampling adequacy. Each of these criteria includes a number of sub-criteria.

The sample of the study consists of nine papers of teacher-made test of English grammar given to the first year college students for five consecutive years.

The researcher has constructed a checklist to include the adopted criteria and the developed sub-criteria. It has been exposed to a jury of experts in order to achieve its face validity. Then the included test papers have been evaluated according to the constructed checklist.

Results of the study indicate that evaluated test papers manifest a lot of defects concerning: layout, accuracy, validity, objectivity, reliability, practicality and sampling adequacy.

II.3.3 Discussion of the Previous Studies (Second Group)

The two previous studies included in the second group aim at evaluating test papers given to college students. The first study, i.e. Jameel (1982) has constructed two questionnaires for collecting the necessary data concerning the ways and systems of the studied examinations. While the second study, i.e. Faisal (2002) has evaluated the intended achievement tests in the light of some criteria and sub-criteria of a good test, and the objectives of the course of study. The present study is also going to evaluate the M.A. achievement tests according to the developed objectives of Bloom's cognitive Domain, as well as the other six criteria, namely: validity, comprehensiveness, objectivity, accuracy, practicality, and layout.

The sample of Jameel (1982) includes thirty-five college instructors and forty-five students, while the sample of Faisal (2002) includes nine test papers. The sample of this study includes twenty-two test papers with 112 questions.

The results of Jameel (1982) state that 79% of college instructors prefer preparing a sample of tests in advance and choosing the most comprehensive, scientific and appropriate ones to their students in the appropriate time of examination. Moreover, college students face many examination problems and prefer having objective tests. The results of Faisal (2002) indicate that college test papers manifest defects concerning: layout, accuracy, validity, objectivity, reliability, practicality and sampling adequacy.

Section Three Procedures

III.1 Introductory Note

This section presents a description of the procedural steps adopted to achieve the aims of this study. It displays the selected sample of M.A. tests, instrument of the study, and the technique of evaluating the included tests.

111.2 Sample Description

The sample of this study implies the written achievement tests of "New Trends to Teaching EFL" subject which are given to M.A. students at the end of the course examinations (first attempts) during the seven consecutive years from 1999-2000 up to 2004-2005. This sample consists of twenty-two test-papers obtained from Baghdad University as well as Al-Mustansiriyah University. The test-papers include 112 questions as shown in Table (2).

From Table (2) different number of questions are involved in each copy of the twenty-two test- papers of the four colleges of education during the seven academic years mentioned above. The total numbers of the questions presented by the College of Education /Ibin Rushed and that for Women are thirty and thirty-one respectively. Only eleven questions are subjected by the two copies of the College of Education during the two identified years and forty questions are included in the seven copies of the College of Basic Education.

Table (2) : Number of Questions of the Twenty- Two Test Papers of the Four Colleges of Education

Baghdad University			Al-Mustansiriyah University		
Year	Education Ibn-Rushd	Education for Women	Education	Basic Education	Total No. of Questions
	No. of Qs *	No. of Qs	No. of Qs	No. of Qs	No. of Qs
1999-2000	4	4	- ** **	6	14
2000-2001	4	5	-	7	16
2001-2002	5	4	-	6	15
2002-2003	4	4	4	7	19
2003-2004	4	5	7	7	23
2004-2005	4	4	-	7	15
2005-2006	5	5	-	-	10
Total	30	31	11	40	112

III.3 Instrument of the Study

A specific checklist is constructed for the purpose of evaluating the selected sample of test-papers. The checklist includes the following major criteria:

1. The cognitive educational objectives of Bloom's Taxonomy which are divided into the following six levels: knowledge, comprehension, application, analysis, synthesis, and evaluation (Bloom, 1956, Huitt, 2004, and Atherton, 2005).
- 2- Validity
 - a- The test provides through its type and content, situations in which the testees exhibit the specific behaviour described in the objectives of the presented course.

* Qs means Questions

** The mark means that there is no M.A. study in some years at those colleges.

- b- The test includes questions that measure the degree of progress of each student towards the intended goals.
- 3. Comprehensiveness:
 - a- The test adequately samples the syllabus items of the intended course of study.
 - b- The percentage of the given questions should correspond well with the proportional emphasis given to the course objectives.
- 4. Objectivity: It is determined by the questions type used, i.e. either subjective or objective.
- 5. Accuracy and Clarity of Wording: The test-paper is carefully planned and worded if each of its questions:
 - a- is preceded by clear instructions.
 - b- refers to a specific material (the independency of a question from other questions).
 - c- has only one possible interpretation.
 - d- is free from grammatical, stylistic, spelling and punctuation errors.
- 6. Practicality and Scorability: The test is practical within the means of financial limitations, time constraints (be specified by time allotted), ease of administration, scoring (has clear distribution of marks), and interpretation.
- 7. Layout and Typography: The test-paper should have attractive appearance, clear printing and format, and enough spaces among its questions and items.

The constructed checklist has been exposed to a jury of twenty specialists. They are asked to verify its face validity by responding to the following question: "are the criteria included in the checklist appropriate to evaluate the efficiency of the M.A. achievement tests?" All the jurors have answered with "yes" as shown below:

Yes: moderate (2 jurors)

Yes: good (10 jurors)

Yes: very good (8 jurors)

This means that the adopted checklist is highly appropriate to undertake the evaluative task of this study. The agreement of eighteen jurors (who represent 90% of the total number of their group) ranges between "good" and "very good" valid, whereas the agreement of only two jurors (who represent 10% of the whole group) takes the moderate level of validity.

III.4 Techniques of Evaluation

In order to achieve the aims of this study, the sample of 112 questions included in the selected twenty- two test papers, the objectives of the intended subject, as well as the proposed criteria are exposed to a jury of fifty college faculties. Those jurors have the academic degrees of full professor or assistant professor of specialization in "Methodology of Teaching English", "Applied linguistics", "Education", and "Educational Psychology".

The specialists involved are first asked to identify each of the exposed questions according to the six levels of cognitive objectives which are put into separate columns in front of the questions. A juror is required to read each question carefully and then tick the appropriate column in front of that question.

After identifying the position of each question in terms of the six cognitive objectives, the juror is asked to read the instructional objectives of the intended subject and evaluate each of the test-papers according to the other six criteria.

Section Four

Analysis of Data and Discussion of Results

IV.1 Introductory Note

The whole responses of fifty specialists on the 112 exposed questions have been collected and calculated into percentages according to the six cognitive levels and other six criteria included in the adopted checklist as follows:

IV.2 The Cognitive Levels

The percentages of the jurors' responses on the checklist items have been calculated. They indicate that the 112 questions included within the M.A. achievement tests are unequally spread over the various levels of cognitive objectives, namely: knowledge, comprehension, application, analysis, synthesis and evaluation.

The first three levels, i.e. knowledge, comprehension, and application get the highest proportions which are; 30%, 23%, and 22%, respectively. The last three levels, i.e. analysis, formulation, and evaluation get the lowest proportions which are; 11%, 8%, and 6%, respectively.

Table (3) : Percentages of Jurors' Responses on the Cognitive Levels of the Analysed Questions

Level	Knowledge %	Comprehension %	Application %	Analysis %	Synthesis %	Evaluation %
Percentage	30	23	22	11	8	6

From Table (3), the low cognitive levels (knowledge, comprehension, and application) have been mostly represented by the analysed questions. This result reflects the deficiency of the M.A. final achievement tests in measuring testees' high levels of thinking. This shows neither the anticipated picture, nor the required efficiency of the

examined questions. The final achievement tests, especially those of higher studies should be mostly related to the highest levels of testees' intellectual abilities, such as analysis, synthesis, and evaluation. In other words, M.A. achievement tests should be relevant to the extent they elicit the desired types of mental processes. They should also be appropriate to the extent that encourages the development of desirable study habits on the part of M.A. students.

IV.3 Validity

As far as "validity" is concerned, the highest percentages of specialists' responses are registered by the third column "moderately valid" and the second column "valid to some extent", whereas the columns "high valid" and "not valid" have registered the least percentages of responses.

From Table (4) most of the M.A. tests are valid but in different levels, as follows:

- a- The tests validity of Baghdad University ranges between "valid to some extent" and "not valid", whereas "highly valid" level has not been occupied by any test.
- b- The tests validity of Al-Mustansiriyah University is distributed between "highly valid", "valid to some extent" and "moderate valid". No test comes at "not valid" level which has not registered any percentage.
- c- The highest percentages of "highly valid" and "valid to some extent" which are sixteen and forty are recorded by the College of Basic Education / Al-Mustansiriyah University.
- d- The highest percentage of "moderate valid" which is sixty eight is registered by the College of Education for Women/ Baghdad University.

- e- By and large, all the involved tests are moderately valid since the highest percentage occupied by each of them falls at "moderately valid" level, which means that M.A. achievement tests moderately measure what they purport to measure.

Table (4) : Validity of the M.A. Final Achievement Tests

	College Name	Response Level		Highly Valid		Valid to Some Extent		Moderately Valid		Not valid	
		No. of Response	%	No. of Response	%	No. of Response	%	No. of Response	%	No. of Response	%
Baghdad University	College of Education-Ibin Rushd	-	-	12	24	28	56	10	20		
	College of Education for Women	-	-	9	18	34	68	7	14		
Al-Mustansiriyah University	College of Education	7	14	18	36	25	50	-	-		
	College of Basic Education	8	16	20	40	22	44	-	-		

IV.3 Comprehensiveness

As far as the criterion of "comprehensiveness" is concerned, Table (5) shows that all the evaluated M.A. tests are fairly comprehensive, but in different degrees as outlined below:

- a- The highest percentages of "moderately", i.e. 74 and 64 have been respectively registered by the tests of the College of Education-Ibin Rushd and College of Education for Women of Baghdad University.
- b- The highest percentages of "comprehensive to some extent" which are 44 and 60 are respectively occupied by the tests of the College of Education and College of Basic Education at AL-Mustansiriya University.
- c- No test occupied the level "not comprehensive".

Table (5) : Comprehensiveness of the M.A. Final Achievement Tests

	Response Level College Name	Highly Valid		Valid to Some Extent		Moderately Valid		Not valid	
		No. of Response	%	No. of Response	%	No. of Response	%	No. of Response	%
Baghdad University	College of Education-Ibin Rushd	-	-	13	26	37	74	-	-
	College of Education for Women	-	-	18	36	32	67	-	-
Al-Mustansiriyah University	College of Education	5	10	30	60	15	60	-	-
	College of Basic Education	-	-	22	44	28	36	-	-

Similarly, no test occupied the level "highly comprehensive" except for the 10% which is occupied by the tests of the College of Basic Education / Al-Mustansiriyah University. This means that the M.A. achievement tests are fairly based on a representative sampling of the presented contents and hence, the percentage of the given questions does not correspond well with the objectives of the given course.

IV.4 Objectivity

Concerning "objectivity", it is quite clear that M.A. achievement tests are fully subjective. The fifty jurors identify that all the evaluated samples of 112 questions (included within the twenty-two test-papers) are 100 percent subjective.

This high level of subjectivity may be due to the nature of the tests or questions given to the M.A. students which should fundamentally require them to explain, analyse, discuss and appreciate some ideas. They may also require them to solve problems, or deal with problematic situations, present some conclusions, and write using their own assumptions and expressions about any subjected question.

Subjective tests can tap high levels of reasoning such as, required inference, organization of ideas, comparison and contrast. Morris (1961: 33) believes that the main advantage of subjective tests is that they are more just to the learner than objective tests because they give a better opportunity to display his/her capacity. Heaton (1975: 11) adds that all tests are constructed subjectively by the tester. He decides which areas of the material to test, how to test those particular areas, and what kind of items to use for this purpose. Thus, it is only the scoring of a test that can be described as objective. Therefore, objectivity is a relative quality and there are no completely objective tests.

IV.5 Accuracy and Clarity of Wording

The collected data indicate that the extent of accuracy and clarity of the M.A. achievement tests ranges from "accurate to some extent" to "not accurate". Table (7) reflects the following results:

- a- None of the investigated tests are worded clearly and accurately to a high extent.
- b- The highest percentages that come under the level "accurate and clear to some extent", i.e. 40 and 38 are, respectively registered by the College of Basic Education and the College of Education at Al-Mustansiriyah University.
- c- The highest percentages which are registered by the M.A. achievement tests for each of the four involved colleges come under the column "fairly accurate and clear". Each of the College of Education / Ibin Rushd, and College of Education for Women at Baghdad University has recorded 68%. The College of Basic Education and the College of Education at Al-Mustansiriyah University, have recorded 58% and 56%, respectively.

d- The lowest percentages come under the column "not accurate and clear".

The above-mentioned results illustrate that the investigated tests are moderately satisfactory in their wording. In other words, the questions posed to M.A. students do not refer to specific material and have many possible interpretations and those questions are not free from grammatical, stylistic, spelling and punctuation errors. These defects will be negatively reflected into testees' responses. Having more than one possible interpretation, the question will mislead the testees and consequently reduce the efficiency of their achievement.

The obtained unpopular level of accuracy and clarity of the evaluated tests could be contributed to their subjectivity (see IV.4). Fernandes and Issa (1972: 7) believe that poor wording of questions, ambiguous statements, and failure to give all pertinent information necessary for students to choose, belong to low objectivity of tests.

Table (7) : Accuracy and Clarity of Wording of M.A. Tests

	College Name	Response Level		Highly Valid		Valid to Some Extent		Moderately Valid		Not valid	
		No. of Response	%	No. of Response	%	No. of Response	%	No. of Response	%	No. of Response	%
Baghdad University	College of Education-Ibin Rushd	-	-	11	22	34	68	5	10		
	College of Education for Women	-	-	13	26	34	68	3	6		
Al-Mustansiriyah University	College of Education	-	-	19	38	29	58	2	4		
	College of Basic Education	-	-	20	40	28	56	2	4		

IV.6 Practicality and Scorability

Analyzing the collected data indicates that the percentages recorded by the M.A. tests concerning "practicality and scorability" fall under the last three levels, i.e. "practical and scorable to some extent", "fairly practical and scorable", and "not practical and not scorable". Table (8) shows the following results:

- a- The highest percentages got by the twenty-two tests fall under the two columns "practicable and scorable to some extent" and "fairly practical and scorable". This means that the evaluated tests are relatively equal in their moderate level of practicality and scorability.
- b- The first level, "highly practical and scorable" has not been occupied by any of the investigated tests.
- c- The fourth column "not practical and not scorable" has registered some percentages of specialists' points of view. Only 20% is recorded by the College of Education /Ibin Rusd and 12% is recorded by the College of Education For Women. Both of these colleges belong to Baghdad University. Moreover, only 10% and 12% are respectively, recorded by the College Education and College of Basic Education which belong to AL-Mustansiriyah University.

This unsatisfactory level of M.A. tests concerning practicality and scorability means that the time allotted to do these tests is inconsistent with the number of the involved questions. It also means that the marks are not specified equally for the posed questions; hence they couldn't be easily scored or interpreted.

Table (8) : Practicality and Scorability of M.A. Tests

	Response Level College Name	Highly Valid		Valid to Some Extent		Moderately Valid		Not valid	
		No. of Response	%	No. of Response	%	No. of Response	%	No. of Response	%
Baghdad University	College of Education-Ibin Rushd	-	-	19	38	21	42	10	20
	College of Education for Women	-	-	20	40	24	48	6	12
Al-Mustansiriyah University	College of Education	-	-	24	48	21	42	5	10
	College of Basic Education	-	-	24	48	20	40	6	12

IV.7 Layout and Typography

As far as the technical make-up of the M.A. tests is concerned, the percentages of specialists' points of view as shown in Table (9) point out the following:

- a- None of the investigated tests come under the column "highly satisfactory"
- b- The highest percentages occupied by each of the College of Education for Women, College of Education, and College of Basic Education fall under the column "fairly satisfactory". These percentages are 46, 50, and 50, respectively.
- c- The highest percentage recorded by the College of Education /Ibin Rushd which is 48 comes under the fourth column, i.e."unsatisfactory"

Table (9) : Layout and Typography of M.A. Tests

	College Name	Response Level		Highly Valid		Valid to Some Extent		Moderately Valid		Not valid	
		No. of Response	%	No. of Response	%	No. of Response	%	No. of Response	%	No. of Response	%
Baghdad University	College of Education-Ibin Rushd	-	-	5	10	21	42	24	48		
	College of Education for Women	-	-	7	14	23	46	20	40		
Al-Mustansiriyah University	College of Education	-	-	13	26	25	50	12	24		
	College of Basic Education	-	-	15	30	25	50	10	20		

- d- The lowest percentages registered by each of the College of Education / Ibin Rushd, and College of Education for Women come under the column "to some extent satisfactory". These percentages are, 10 and 14, respectively.
- e- The lowest percentages registered by each of the College of Education, and College of Basic Education come under the column "unsatisfactory ". These percentages are, 24 and 20, respectively.

The obtained results indicate that M.A. achievement tests suffer from a clear shortage in their technical aspects. This shortage may be due to the features that characterize the evaluated tests. Having unattractive appearance, unclear printing and format, insufficient spaces among their questions and items, unequal distribution of marks on the questions included, play one role or another in minimizing the technical efficiency of the tests.

Section Five

Conclusions and Recommendations

V.1 Conclusions

As a result of evaluating the efficiency of the tests under this study and in the light of the empirical evidences reached at, some points are concluded as follows:

- a- the overall efficiency of M.A. final achievement tests has been of unsatisfactory level and with approximate degrees of deficiency across the various criteria of evaluation. This means that they have failed to satisfy any of the adopted criteria.
- b- By and large, the investigated tests of the four involved colleges across the seven previous academic years, have occupied similar levels of efficiency which range between the second and the third one.
- c- College faculties follow a traditional way in constructing their tests and there is not any serious attempt of making some changes or innovating that traditional way. This could be easily concluded from the similar types of questions that are posed to the M.A. students during the seven consequent years.
- d- M.A. final achievement tests involve more or less exclusive types of questions which are unable to elicit the desired types of mental processes. These tests mostly demand rote learning and consequently encourage M.A. students to build up their study habits accordingly, i.e. studying without being able to analyse, apply, appreciate, or evaluate what they read.
- e- College faculties lack experience, technical training, and clarity of vision concerning the processes of evaluation and tests construction. These essential prerequisites should have been already gained through intensive practice of professional qualification.

V.2 Recommendations

As far as the above-mentioned conclusions are concerned, some recommendations are put forward:

- a. College faculties should well realize the characteristics, principles, and sensitivity of the process of constructing their tests in general, and M.A. achievement tests in particular.
- b. M.A. achievement tests should concentrate on the high cognitive levels of testees' thinking such as, analysis, formulation, and evaluation, so that students' performance in these tests could reflect the extent to which they achieve the intended instructional objectives.
- c. Some conferences and symposiums should be occasionally held at Iraqi universities to discuss and exchange the various points of views about tests construction. Such kinds of discussions and exchanges of views will keep college instructors in frequent contact with recent innovations, and issues of periodicals concerning tests and tests construction.
- d. A successful college tester is the one who looks at the process of test construction as an art which should be related as much as possible to contents and behaviours of the presented course or material.
- e. A careful initial planning, specifying the aims clearly and considering the criteria of a good test make it possible to avoid pitfalls, assure more efficient testing procedures, and results in a better end product.
- f. Postgraduate EFL students should be trained in how to perform well on different types of tests that measure their abilities, skills, and achievements.
- g. An M.A. test designer should have clear and carefully planned scoring scheme through which the total mark is distributed among the various questions and items included within the given tests. This will reduce the independency of the subjective judgment and the individual impression of the evaluator.

References

1. Atherton, J.S. (2005) Learning and Teaching: Bloom's Taxonomy, UK. London
2. Bloom, B.S. (2005) Taxonomy of Educational Objectives: The Cognitive Domain, New York, David Mckay.
3. Ebel, Robert L. (1972) Essentials of Educational Achievement, Englewood Cliffs, Prentice-Hill Inc.
4. Everitt, A. and Hardiker, P. (1996) Evaluating for Good Practice, London: Macmillan.
5. Faisal, Wafaa M. (2002) Evaluation of Language Tests of Iraqi Universities, Departments of English, (Unpublished Ph.D. Dissertation), Baghdad University.
6. Frenands, H.J. and Mustafa M. Issa (1972) Evaluation of Annual English Classroom Tests in Grade V of the Primary Schools, Iraq, University of Baghdad, Educational Research Center.
7. Huitt, W. (2004) Bloom et al's Taxonomy of the Cognitive Domain. Educational Psychology Interactive. Valdosta, GA: Valdosta State University.
8. Jameel, M. (1982) Studying the Status of Geometric Examinations, (unpublished M.A. Thesis), Baghdad University.
9. Jeffs and Smith, M.K. (2005) Informal Education. Conversation, Democracy and Learning 3e, Nottingham: Educational Heretics Press.
10. Lyman, Howard B. (1971) Test Score and What They Mean, 2nd edition, New Jersey: Prentice-Hall, Inc. Englewood Cliffs.
11. Morris, Norman (1961) "A Historians' View of Examination in Stephen Wiseman (ed.)", Examinations and English Education, London: Manchester University Press.
12. Rubin, F. (1995) A Basic Guide to Evaluation for Development Workers, Oxford, Oxfam.

- 13.Scott, L. Howell and Mary Hricko (2005) Online Assessment: Case Studies from Higher Education. Idea Group, Inc.
- 14.Smith, Mark K. (2006) "Evaluation" in the Encyclopedia of Informal Education, [www.infed.org/ biblio/b.eval.htm](http://www.infed.org/biblio/b.eval.htm).
- 15.Stiggins, R. J. (2003) Students-Centered Classroom Assessment, New York: Merrill.
- 16.Wilhelms, F. T. (1967) Evaluation as Feedback and Guide, Yearbook, Washington, Association for Supervision and Curriculum Development.