

# ***THE EFFECT OF FORMATIVE ASSESSMENT ON THE PUPILS' ACHIEVEMENT IN ENGLISH***

**A DISSERTATION SUBMITTED TO**

**THE COUNCIL OF THE COLLEGE OF EDUCATION / IBN  
RUSHD – UNIVERSITY OF BAGHDAD IN PARTIAL  
FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE  
OF DOCTOR OF PHILOSOPHY IN MEHTODS OF  
TEACHING ENGLISH AS A FOREIGN LANGUAGE**

**BY**

***RIDHA GHANIM DAKHIL AL-SAEDI***

**SUPERVISED BY**

**ASSISTANT PROFESSOR  
DR. ABDULLATIF AL-JUMAILY**

**PROFESSOR  
DR. ABDULLAH H. AL-MUSAWI**

**1425 A.H.**

**2004 A.D.**

# أثر التقويم التكويني في تحصيل الطلبة في اللغة الإنكليزية

أطروحة

مقدمة إلى مجلس كلية التربية (ابن رشد) في جامعة بغداد

وهي جزء من متطلبات نيل درجة دكتوراه فلسفة في

طرائق تدريس اللغة الإنكليزية

من

رضا غانم داخل الساعدي

بإشراف

الأستاذ المساعد الدكتور

عبد اللطيف علوان الجميلي

الأستاذ الدكتور

عبد الله حسن الموسوي

٢٠٠٤ م

١٤٢٥ هـ

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

((قُلْ لَنْ يُصِيبَنَا إِلَّا مَا كَتَبَ اللَّهُ لَنَا

هُوَ مَوْلَانَا وَعَلَى اللَّهِ فَلْيَتَوَكَّلِ الْمُؤْمِنُونَ \*))

صَدَقَ اللَّهُ الْعَظِيمُ

سورة التوبة - الآية (٥١)

# *Dedication*

*To :*

- My Mother and Brother's Soul*
- My Father*
- My Wife , My Sons and Daughters*

بسم الله الرحمن الرحيم  
إقرار الخبير اللغوي

أشهد بأنني قد قرأت هذه الأطروحة الموسومة (أثر الاختبار التكويني على  
تحصيل الطلبة في اللغة الإنكليزية) التي قدمها الطالب (رضا غانم داخل  
الساعدي) إلى كلية التربية / ابن رشد / جامعة بغداد ، كجزء من متطلبات نيل  
درجة دكتوراه فلسفة في التربية (طرائق تدريس اللغة الإنكليزية لغة أجنبية) وقد  
وجدتها صالحة من الناحية اللغوية .

الأستاذ المساعد الدكتور

مؤيد محمد سعيد

بسم الله الرحمن الرحيم  
إقرار الخبير العلمي

أشهد بأنني قد قرأت هذه الأطروحة الموسومة (أثر الاختبار التكويني على  
تحصيل الطلبة في اللغة الإنكليزية) التي قدمها الطالب (رضا غانم داخل  
الساعدي) إلى كلية التربية / ابن رشد / جامعة بغداد ، كجزء من متطلبات نيل  
درجة دكتوراه فلسفة في التربية (طرائق تدريس اللغة الإنكليزية لغة أجنبية) وقد  
وجدتها صالحة من الناحية العلمية .

الأستاذ المساعد الدكتورة  
ناهدة طه مجيد الناصري

## الخلاصة

يؤدي التقويم دوراً أساسياً في عملية التعلم و التعليم لما له من أهمية كبيرة في نجاح هذه العملية والتي ستظل من دونه مجرد عملية لا نعرف ما تحقق من الأهداف التربوية الخاصة أو العامة . و لا نعرف مدى التطور الحاصل في تعلم الطلاب و مدى استجابتهم ، والذي من خلاله أيضاً يعرف المعلم مدى نجاح طريقته في التدريس و كذلك ولي أمر الطالب الذي يستطيع الوقوف على مستوى أبنائه من خلال تقويم أدائهم في المدرسة ، ناهيك عن دور التقويم في انتقال الطلاب من مرحلة إلى أخرى و معرفة الفروق الفردية بينهم .

لذا عمد الباحث التعرف على أثر نوع من أنواع التقويم في تحصيل الطلبة في مادة اللغة الإنكليزية من خلال الكشف التجريبي الذي قام به خلال هذه الدراسة ، و هذا النوع هو التقويم التكويني .

تقتصر الدراسة الحالية على طلاب الصف الثاني المتوسط في محافظة بغداد للعام الدراسي ٢٠٠١-٢٠٠٢ و تحدد الدراسة كذلك على فعاليات التركيب ، المفردات ، الاستيعاب القرائي و الإنشاء في كتاب اللغة الإنكليزية لهذه المرحلة .

و لتحقيق هدف البحث صمم الباحث تجربة تتضمن مجموعتين من الطلبة تجريبية و أخرى ضابطة . و تم اختيار هاتين المجموعتين من شعب الصف الثاني في متوسطة الارتقاء للبنين و لكل مجموعة (٣٥) طالباً و قد اختيرت إحداهما لتكون مجموعة تجريبية تم تطبيق التقويم التكويني عليها و الأخرى ضابطة يتم تقويمها بالتقويم الختامي . كوفئ أفراد المجموعتين إحصائياً في المتغيرات : التحصيل الدراسي للأبوين ، العمر الزمني ، التحصيل في مادة اللغة الإنكليزية في السنة السابقة و في الشهر الأول من السنة الحالية .

قام الباحث نفسه بتدريس مجموعتي البحث ، و هذه المادة هي الوحدات من ٦-١٣ في الكتاب المقرر و قد استغرقت التجربة (٨) أسابيع .

قدم الباحث اختباراً تحصيلياً مشتملاً على (٤٨) فقرة في ضوء تحليل دقيق لمحتويات الوحدات الثمان (مادة البحث) ، تم عرض الاختبار على مجموعة من

الخبراء للتأكد من صدقه الظاهري . و تم احتساب ثبات الاختبار باستخدام طريقة إعادة الاختبار ، كما أجري تحليل فقرات الاختبار لتحديد مستوى الصعوبة و قوة التمييز ، و طبق الباحث الاختبار على أفراد الدراسة في نهاية التجربة .

و قد استخدمت معادلة الاختبار التائي لعينتين مستقلتين لاختبار دلالة الفرق بين متوسطي درجات مجموعتي البحث . كما استخدمت المعادلة نفسها لمعرفة دلالة الفروق بين متوسطي مجموعتي البحث في التركيب ، المفردات ، الاستيعاب القرائي ، و الإنشاء و قد وجد أن الفرق كان ذا دلالة إحصائية لصالح المجموعة التجريبية .

استنتج الباحث أن استخدام التقويم التكويني يزيد من تحصيل الطلبة و يعزز التعلم . و على ضوء نتائج البحث تمت صياغة عدد من التوصيات و المقترحات .



## **Acknowledgements**

The completion of this study would not have at all been possible without the help and support of many people to whom acknowledgement is made with great thanks .

I would like to express my sincere thanks and deep gratitude to my supervisors , Dr. Abdullatif Al-Jumaily and Dr. Abdullah Al-Musawi for their invaluable suggestions , support and encouragement throughout the various stages of the study .

Special gratitude is also addressed to the members of the seminar committee , especially , Prof.Ayif Habeeb , Dr.Ibrahim Al-Shibly , Dr.Nahida Taha , Dr.Istiqlal Al-Marsoomi , Dr.Abdul-Jabbar Darwesh, and , Mrs.Lamya Al-Ani .for their constructive suggestions and to the jury experts for their useful comments . My thanks are extended to Dr.Dhuha Atallah , and Dr. Sa'di Jasim for their brotherly help and support .

I would like to express my gratitude to Mr.Hassan Tu'ma , the headmaster of Al-Irtiqa' Intermediate School , Mr.Ibrahim Esa the teacher of English in the school for their assistance in the application of the test . The researcher can only hope that when reading the present study , all those who in any way contributed to it will feel their advice has not been wasted . He also hopes they will forgive him for not referring to them individually .

I am also thankful to my classmates , Madiha Saif , In'am Yousif and Attallah Abed Raba for their brotherly assistance throughout the preparation of this study .

My gratitude , love and acknowledgement are sincerely devoted to my family members who offered great help and continuous encouragement .

## Chapter One

### *Introduction*

#### 1.1 The Problem

The term education is the process of causing behavioural changes . Educational objectives are essentially changes in the human beings , i.e. , the objectives aimed at are to produce such desirable changes. If we notice a desirable change in the behaviour of the human beings , this means that learning has taken place and the instruction process has followed the right way .

The teaching / learning process is a decisive step in the whole educational setting . This process includes many components to be effective and useful .

One of these components is teaching . Teaching is the organization and guidance of learning . It also involves the ability to react to a different and complex factor , such as the demands of certain curriculum to the needs of certain pupils ( Applebee , 1988 : 213 ) . The teacher should have certain characteristics to be successful in this filed . He should know much about the learners' needs and their interests , the best methods of motivating them and evaluating their achievement (Bronghton et al ., 1981:37).

The second component of the teaching / learning process is learning . Foreign language learning is complex, so no single development can solve all its problems . In order to learn a language , the learner should practice and participate . Thus , the learner will be eager and interested in doing things that lead to

learning ( Lado 1972 : 41 ) . Learning a language is a behaviour , and this behaviour can be acquired by convincing the learner to perform in the language .

The third component which can be considered as the cornerstone of the teaching / learning process is the appraising of this behaviour or the changes that might take place during this process ( Rivers . 1972 : 15 ) .

The last decades have witnessed many strategies in the various aspects of the teaching / learning process . As a result , teachers have a wide variety of teaching methods and techniques at their disposal . To select from these confusing and contradictory types of methods , the teacher has to evaluate wisely the comparative effectiveness of various innovations in the classroom environment .

Appraising of these changes or objectives is the job of assessment , which is the process for determining the degree to which these changes are actually happening .

In addition , it reveals the strengths and weaknesses of the plans . As a result of assessment , it is possible to note in what aspects the programme is effective and in what aspects it needs improvement .

Scannel and Tracy (1975 : 5) summarize the teaching-learning process in the following figure :

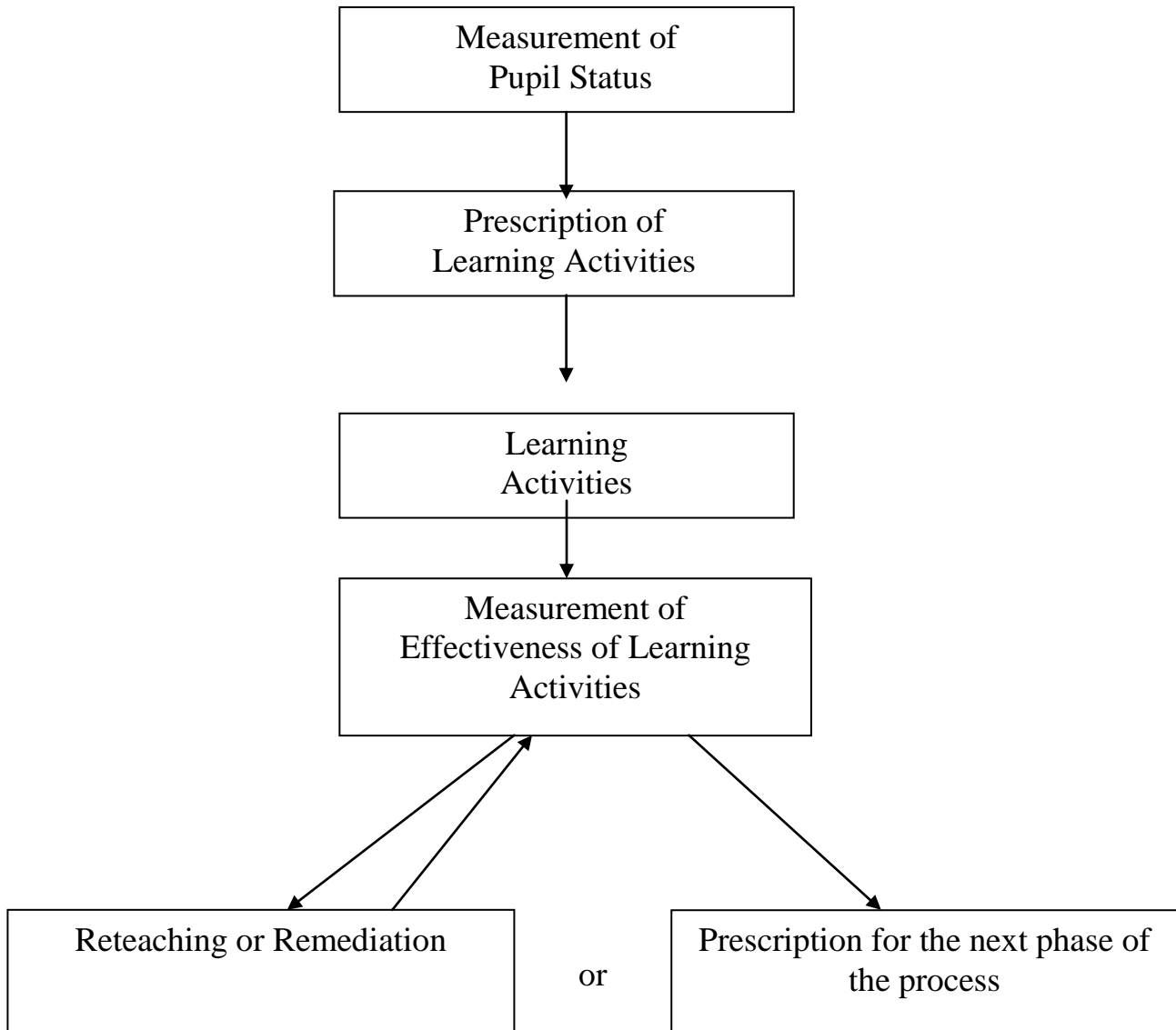


Figure (1)  
The Continuous Teaching – Learning Process

The intelligent application of educational assessment devices can improve the quality of Education . Educational assessment focuses not only on the nature of the learner , but also on what is to be learned and how . The specialists are concerned with the strengths and weaknesses of the learner , the effectiveness of the instructional materials , and curriculum , i.e. educational assessment is diagnostic in intent .

Callahan ( 1983 : 220 ) indicates that assessment gives data which serve as a basis for monitoring the gradual development of the learning process in the mind of the learner . It is also used to guide teachers and to motivate learning .

The findings of the assessment give those who are concerned with this field an enlightenment about the status of their pupils which helps them make a judgment which in turn leads to decision-making .

It is useful and necessary to pinpoint the terms , measurement , evaluation and assessment which wear new meanings when used by specialists , while they are interchangeable when they are commonly used .

Payne ( 1974 : 8 ) states that " *measurement is concerned with the systematic collection , quantities , and ordering of information* " . This means that measurement refers to the quantitative description of pupils ( i.e. test scores ) .

Gronlund ( 1965:6 ) defines evaluation as a systematic process of determining the extent to which educational objectives are achieved by pupils . This definition assumes that evaluation

implies both qualitative and quantitative descriptions of pupils' behaviour plus value judgment . In other words , measurement does not imply quantitative description nor does it imply judgments concerning the value of behaviour measured .

Now one can conclude that within the umbrella category "*assessment*" , evaluation is a much more comprehensive and inclusive term than measurement . Assessment concerns itself with the totality of the educational settings , i.e., it subsumes measurement evaluation .

Scannell and Tracy ( 1975:6 ) assure that evaluation gives feedback to the teachers , pupils , and parents about each pupil's progress and the teacher may change or modify the plan accordingly depending on the information obtained by the assessment . Consequently , the teacher can decide whether the pupils is ready for the next step in the school programme or not .

The achievement of goals is carried out through two major types of assessment , namely : formative and summative assessment .

According to Carol ( 1997 : 2 ) , formative assessment refers to the process of gathering and integrating information about learners during instruction to help us understand these students and describe them . Summative assessment , on the other hand , refers to the appraisal of pupils' achievement after they have finished the unit or the course .

To Popham ( 1993 : 19 ) formative assessment refers to the appraisals of quantity focused on instructional programme that are

still capable of being modified , while summative assessment refers to the appraisals of quality focused on completed instructional programme .

Thus Callahan ( 1983 : 220 ) holds that formative assessment happens as the unit , course , or sequence progresses .It is used as a tool of diagnosis , as a basis for remediation and as an indication to students' strengths and weaknesses to help teachers build on the students' strengths and correct their weaknesses before the instruction ends , while the summative assessment sums up the results of the instruction and the progress of the group and it takes place after the materials that have been accomplished .

English educators often agree that their appraisal needs are not met by summative evaluation alone . It is fruitless and frustrating to discover at the end of an academic year that the methods used were ineffective and unsuccessful . Teachers need an evaluation of instruction – in – progress to modify or change their methods midway or during teaching / learning process . Formative assessment helps teachers identify weakness of the pupils during each learning unit . It also assists teachers enhance the strong points of their pupils and teaching / learning process. The feedback provided by formative assessment is likely to help pupils become active participants in the teaching / learning progress .

The above stated salubrious effects of the use of formative evaluation in English classrooms are expected outcomes or tentative hypotheses which need to be empirically tested .

As a result of surveying the points of view of a number of supervisors and teachers of English concerning the procedures or strategies followed in Iraqi schools as far as assessment is concerned , it is concluded that no systematic procedure is prescribed or recommended to conduct formative assessment . But when one reviews the Iraqi examination system , one finds that a considerable amount of marks is allotted to daily classwork , homework and quizzes . Therefore , there is some sort of contradiction between what is supposed to be carried out with regard to formative assessment and what is actually applied by the teacher inside the classroom . This contradiction creates a gap between theory and practice which might lead to serious consequences on the pupils' performance in particular and the whole teaching / learning process in general .

The existence of such a gap as well as the scale of its seriousness has not yet been pointed out or investigated . The present study is an attempt to bridge this gap in the literature and provide a foundation to build upon in this respect .

## 1.2 Aims of the Study

This study aims to :

1. examine the effect of formative assessment on the pupils' achievement in English at the second intermediate grade , and
2. identify the quantitative and qualitative differences in pupils' performance when using formative assessment systematically and when using the currently followed ways of assessment .



### 1.3 Hypothesis

It is hypothesized that there is no statistically significant difference between the mean score of the achievement of the formative assessment group and that of the summative assessment group in the post test .

### 1.4 Significance of the Study

Block ( 1971 : 38 ) sees that the flexibility of formative evaluation allows teachers to find out the problematic areas during the teaching / learning process and provides them with feedback to make any necessary modification in the methods used inside the classroom . It also enables teachers to use remedial exercises before these problematic areas have bad effects on students' achievement .

Thorndike ( 1969 : 178 ) states that evaluation is a major aspect of the teacher's job . Its implementation in the proper way is of great importance to ensure taking the right decision concerning pupil's performance and to enable the teacher to have a clear – cut picture of how the learners are progressing , and the extent to which his teaching has been effective .

Formative test is one of the most important tools by which the teacher can appraise the pupils' achievement in order to determine the weak points and use the suitable strategy to remedy these points , and to enhance the strong ones . The aim , therefore , is to help pupils become active participants during the learning unit .

The researcher hopes that the findings of this study will also be of great important to textbook writers and syllabus designers in that they will know which type of assessment is more effective and successful than the other and what solutions they can suggest to be followed in order to improve the learners' performance in particular and to diagnose the weak points that may hinder teaching / learning process in general .

### 1.5 Plan of the Study

To fulfil the aims of the study , an experiment will be conducted . Two groups of pupils will be selected ; a control group and an experimental one . These groups will be chosen from one of the Iraqi intermediate schools . The researcher himself will teach both groups certain units of the prescribed textbook using the same method of teaching. However, testing will be different for both groups; the experimental group ( the formative assessment group ) will be tested informally and continuously at certain intervals . While, the control group (the summative assessment group) will be tested only at the end of experiment . This final test will be administered to both groups simultaneously .

### 1.6 Limits of the Study

1. This study is limited to the male pupils of the second grade , intermediate schools in Baghdad during the academic year (2002-2003) .

2. The material is restricted to oral practice , reading and writing in units (6) to (13) of book 4 (NECI) .

## 1.7 Definition of Basic Terms

### Formative Assessment

It seeks to improve the learning / teaching system by feedback of information from test results which can illustrate the effectiveness of teaching method , or highlight learning difficulties (Page , 1977 : 138) .

The operational definition here is that formative assessment is a form of a criterion-based testing used to diagnose students' errors and misunderstandings for the purpose of identifying instructional strategies appropriate to the students' needs .

### Achievement

- It can be defined as the learning that takes place during a definable course of instruction ( Dwyer , 1982 : 12 ) .
- Good ( 1973 : 7 ) refers to achievement as " *the accomplishment or proficiency of performance in a given skill or body of knowledge* " .
- The operational definition here is that achievement is the knowledge attained or skills developed in the school subject measured by test scores which the pupil sets on the final achievement test which the researcher had prepared for the purpose of this study .

## Feedback

It is the programmed instruction term for the information needed to modify a programme so that failure in responding can be minimized .

It can be used to denote the influence on a unit of behaviour of one's knowledge of the effects of the last unit (Page, 1977:134) .

## Chapter Two

### Theoretical Background

#### 2.1 An Introductory Note

The process of evaluation is essentially the process of determining to what extent the educational objectives are realized by the programme of curriculum and instruction . Hence evaluation is an important device to clarify educational objectives if they are not clarified in the curriculum planning .

Evaluation has an influence on learning . Students , to some extent , are influenced by the type of evaluation made , and the teachers are also influenced in their emphasis in their teaching by the kind of evaluation which they expect to be made .

Tyler (1969 : 134) believes that evaluation plays a great role in the guidance of pupils . It is valuable to know much about the background of the pupils and their achievement of different kinds of objectives in order to know their needs and their capabilities .

Evaluation is considered a basic tool used by teachers and evaluators for appraising students' performance in order to show the extent to which specific goals and general aims of the educational process have been achieved , and then to make any necessary change where it is needed .

Bloom et al. (1971 : V) believe that evaluation should enable teachers to make improvements in their students' learning depending on the information produced by evaluation . This information gives teachers enlightenment about the position of their students and to make a judgement which leads to decision making .

The other functional value of evaluation is that it can be used as a basis for finding out points that need further attention on the part of specialists of this field with particular groups of students and as a basis for planning programmes for the students in the light of their particular progress in the educational programme . (Block , 1971 , Brown , 1976) .

Evaluators agree that the best evaluation should be continuous , comprehensive and consistent with the objectives of the programme . Most evaluators indicate that there are two main kinds of assessment : formative assessment and summative assessment .

## 2.2 Formative Assessment versus Summative Assessment

There are many functional definitions of the formative assessment . Carey and Dick (1981 : 158) refer to formative assessment as "*the process used to obtain data for instruction to be used to increase the efficiency and effectiveness of their instructional materials*".

Popham (1993 : 83) points out that the formative evaluators gather information regarding the worth of aspects of an instructional sequence in order to make the sequence better . Still-malleable instructional sequences that formative evaluation might serve would include :

1. an early version of a set of self – instructional booklets  
or
2. a newly initiated educational programme where the faculty is still trying to devise effective components of the programme .

This means that formative evaluators try to assess such programme in order to draw the attention of the programme developers to how to improve faults in their instruction .

Bloom et al. (1971 : 133) hope that the users of formative evaluation will relate the results of the evaluation to the learning and instructional goals they regard as important and worthwhile .

Summative evaluators , on the other hand , gather information regarding the worth of an overall instructional sequence so that decisions can be made regularly whether to retain or adopt that sequence . This means that the summative evaluation is used at the end of a course and the results of the evaluation will be the basis on which they depend in grading the learners and the instruction may be modified accordingly .

Page et al. (1977 : 138) define formative assessment as " *the process which seeks to improve the learning/ teaching system by feedback of information from test results which can illustrate the effectiveness of teaching methods , or highlight learning difficulties* " .

We can conclude from the above-mentioned definition that feedback is essential and it is considered the cornerstone of this

type of assessment . Hence we shall emphasize the effect of feedback on the teaching / learning process .

According to Sadler ( 1989 : 120 ) formative assessment is initially connected with feedback to both teachers and learners . Teachers use feedback to make programmatic decisions with respect to readiness , diagnosis and remediation . Students on the other hand , use feedback to recognize and reinforce their strength and to improve their weaknesses .

As far as feedback is concerned, Gipps (1994 : 125) warns that grades are not referred to as feedback . Feedback is so when information is used to close the gap between the actual performance and the desired performance .

Feedback is considered the backbone of the teaching / learning process . It includes certain facets such as praise , blame , encouragement and other rewards and punishments that support learning as crucial factors of the instructional process . (Bloom et al. 1976 :203) . This means that feedback enhances correct responses and corrects errors or helps the learner correct such errors .

Coopersmith (1967 : 4f ) avers that feedback gives a chance to individual pupils to express their attitudes of approval or disapproval and indicates the extent to which an individual believes himself to be able , significant , successful and worthy .

Scannell and Tracy ( 1975 :7f ) believe that optimal learning occurs when the learner knows his progress . This means that a pupil cannot go forward effectively without knowing how well he has



achieved the preceding step . This can be carried out by writing comments on the paper returned to the students or by using verbal reaction to their efforts .

The teacher makes use of formative assessment , since it provides the teacher with a chance to modify the method of teaching he followed in time to make any necessary modification to suit the new situation and to solve the problem he / she may encounter during the time of the instruction .

Many evaluators agree that the teacher who uses formative strategy may achieve the following :

- a) Determining what each learner has learned in a particular unit and what the learner still needs to learn .
- b) Providing feedback to both teacher and learner on the aspect of the unit that still needs to be mastered .
- c) Showing the students' learning success and failure in order to take the right steps to go forward in instruction and learning process (Patton , 1987:129 ; Gronlund , 1982:4 ;Bloom et al. , 1971:119) .
- d) Helping the teacher to know how to remedy the learning failure and to find a suitable instructional method .
- e) Helping the students to reinforce what they have mastered .

Glaser (1963 : 519) refers to formative assessment as a criterion-referenced testing , while summative assessment as a norm-referenced testing . He indicates that criterion-referenced measures depend upon an absolute standard of quality while

norm-referenced measures depend upon a relative standard . This means that criterion-referenced testing assesses the degree of competence of a particular student without reference to that of other .

Scannell and Tracy (1975 : 39) summarize the criterion -referenced teaching model and norm – referenced teaching model in the following two figures :

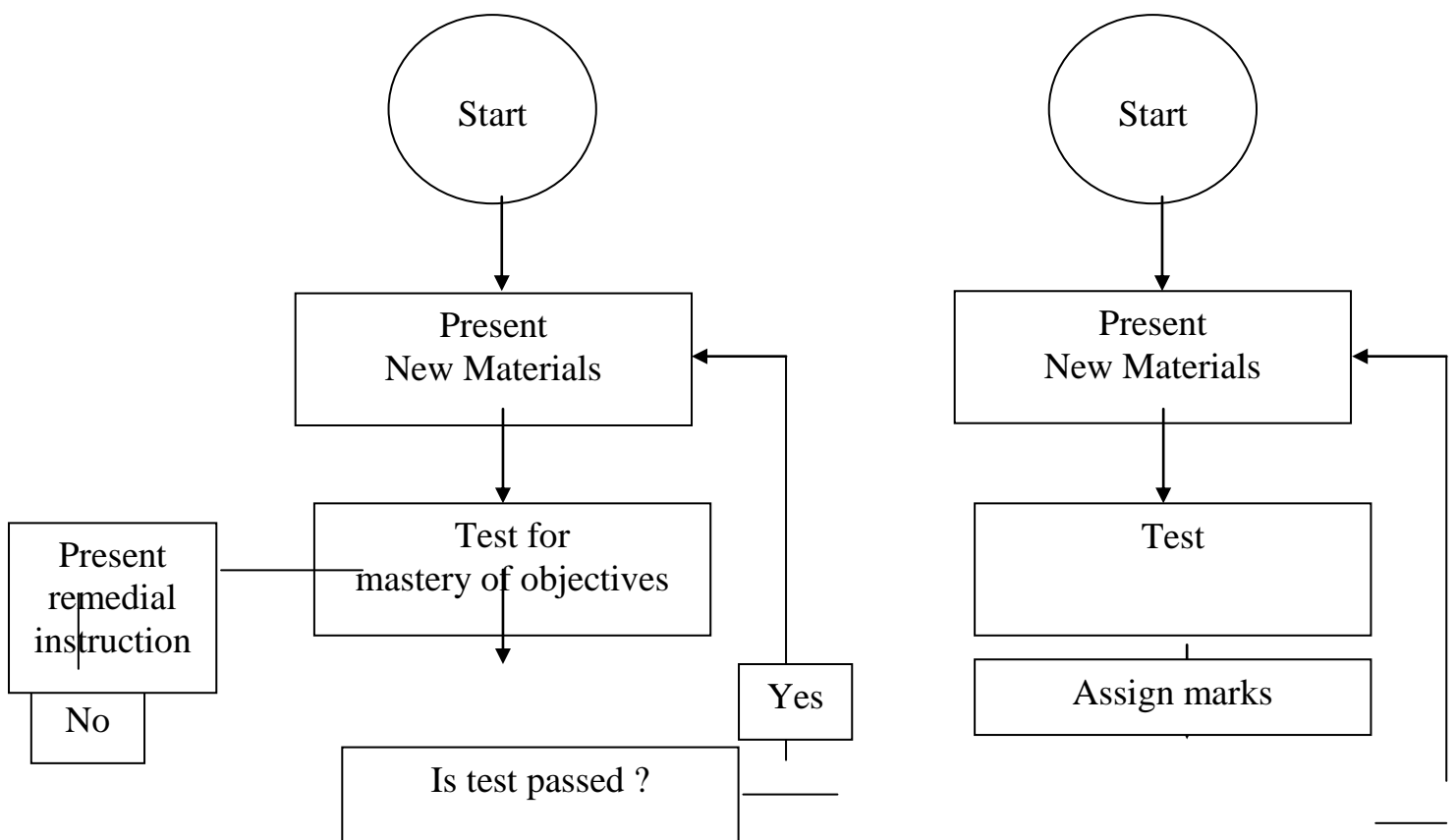


Figure (2)

A criterion-referenced  
teaching model

Figure (3)

A norm-referenced  
teaching model

Ebel (1979 : 11) , on the other hand , differentiates between the criterion–referenced test and the norm–referenced test in four major respects :

1. Criterion – referenced test determines specific objectives achieved by individual pupils , while the other determines the overall knowledge of the subject .
2. Criterion – referenced test indicates pupils' attainment of specific instructional objectives , while norm – referenced test determines a pupil's standing in the group .
3. Criterion – referenced test concentrates on a limited number of objectives , while norm – referenced test covers a large number of subjects .
4. Criterion – referenced test is used to make sure that the pupil learns certain things , while the other is used to indicate a pupil's degree of success in learning .

We can conclude that criterion – referenced test takes care of the individual's growth rather than variations between individuals . Here we have a simple question to be answered by criterion – referenced testing . The question is " *Can X and Y do Z ?* " rather than " *Did X score higher than Y on Z ?* " .

### 2.3 Testing

As far as the teaching / learning process is concerned , testing is regarded as an important step in this process since it is one of the most important ways by which one can measure the behaviour of

persons at a certain time . A test serves two main purposes ; as a guide to the students and a guide to the teachers .

A test helps the pupil reveal weaknesses and reinforce what he knows . It also helps the pupil learn the materials , since it is considered as a learning device . A well – constructed test will be a help for students and to encourage them for self – improvement . A good test is also a guide to the teacher to pinpoint the weaknesses of methods followed in teaching to be modified or changed in order to make the teaching more effective .

A test is an important tool of evaluation . Gronlund (1965 : 11) defines a test as " *a series of tasks which is used to measure a sample of person's behaviour at a given time* ". This means that a test is a systematic way of gathering information on which a tester can evaluate individual performance which, in turn , leads to decisions as far as the teaching / learning process is concerned .

The teacher uses tests for different purposes such as: helping him / her to increase his / her own effectiveness , locating the areas of difficulties encountered by students , evaluating the effectiveness of the syllabus as well as the methods and materials he / she is using . (Scannell and Tracy, 1975 : 21 see also Darwash, 1989 : 8) .

## 2.4 Types of Tests

There are many kinds of tests . A number of test specialists agree that there are two main types of test :

1. Oral test .
2. Written test (teacher – made and specialist - made tests) .

Written tests can be either informal that can be constructed by the classroom teachers , or formal (standardized test) which is designed by specialists in which the examinee is aware that he is being tested for a certain purpose at a particular time . Written tests can be classified into two types according to scoring :

1. Essay test (subjective test) : This type of test is subjectively scored . This means that the results of the test are affected by the opinion of the scorer , and the marks may differ from one scorer to another .
2. Objective test : All tests are constructed subjectively by the tester ; the tester chooses the items to be used , he decides which areas to test and the way by which he can test these particular areas . The test is called objective when the testee scores the same mark even when there are more than one scorer.

#### 2.4.1 Types of Tests according to Function

The question to be answered here is : *Why do we test ?* The answer of this question is the function of each test that the tester aims at . This means that the tests are after the functions they are to serve . These tests are :

1. Mastery Test : This type of test measures the knowledge , facts that all pupils have to acquire . This kind sets a minimum as a standard which all pupils are expected to achieve . Pupils who

make errors should practice till they reach the perfect or near-perfect scores . (Heaton ; 1975:163)

2. Survey Test : The aim of this type is to measure the extent of differences in various pupils' achievement . The total score and the varying degrees of success among pupils are the main points.
3. Diagnostic Test : This type focuses on errors encountered by pupils so that appropriate remedial work can be taken later . (Gronlund: 1977:2 , Payne , 1974:8 , Gronlund: 1965:10) .

#### 2.4.2 Types of Test according to the stages of Instruction

Actually there are three types of test that are typically associated with each stage of instruction . They are :

##### 1- Placement test :

This type of test is given at the beginning of a course of instruction . It can be subdivided into two kinds of testing . The first is readiness pretest in which pupils lack in fundamental skills could be given remediation . The second is placement pretest which sheds light on the student to be placed at the most suitable position in the teaching process . (Heaton ; 1975:168)

##### 2. Formative and diagnostic test :

The aim of an instructional programme is the students' progress made by them during this programme . Formative test is constructed to test the extent to which students have mastered the learning outcomes of , for example , a unit or a segment of

instruction . The purpose is to identify the positive and negative results in order to alter the methods of teaching .

Diagnostic testing , on the other hand , is designed to examine deeply the causes of the learning failure which is not remedied by formative testing .

### 3. Summative test :

This is used at the end of a course of instruction to measure the extent to which the students have achieved the intended outcomes of the instruction to certify mastery or assign grades . (Harrison , 1983 : 7, Gronlund , 1977 : 55 ; Bloom et al. 1971 : 189) .

#### 2.4.3 Types of Test according to content

Oller (1979 : 36-7) illustrates that there are two main types of test as far as content is concerned , they are :

- 1- Discrete – point testing which attempts to assess one bit of a language at a time . This means that language has not been taken as a whole to be tested .
- 2- Integrative testing : This type opposes the previously mentioned type , i.e. , this kind tries to assess many components all at the same time . This means , it puts language skill back together .

## 2.5 Formative Test construction

The most important value of formative evaluation is that it gives an aid to the pupil in learning . Formative evaluation can be accompanied by different materials and procedures by which a learner can remedy the gaps in learning a specific unit .

The other effective use of this type of evaluation is in pacing pupil's learning when the subject is sequential in a course . This means that the next subject depends on the pervious one (i.e.) poor learning of the early unit leads to poor learning of the next unit .

To construct a formative test , the teacher has to take into account the following steps :

- 1- Determination of the objects and the concepts that the student has to acquire of each study unit .
- 2- Specifying the time allotted for the objectives to be achieved .
- 3- Determining the methods and procedures to be followed inside the classroom .
- 4- Choosing the items that can measure the achievement of the aim of instruction .
- 5- Arranging the common mistakes in order to use an appropriate method to remedy such mistakes and to help students get rid of these mistakes during the studying process (Karmel 1977 : 283 ; Bloom et al. 1971 : 119) .

## 2.6 Test qualities

Any one working with tests should understand certain qualities that a good test should have in order to construct and interpret the test results .



## 1- Construct Validity :

Test validation is a continuous process which begins with the design of a test and continues with collecting of evidence to assist the intended interpretations . Bachman and Palmer (1996 : 21) indicate that " *construct validity pertains to the meaningfulness and appropriateness of the interpretations that we make on the basis of test scores* ". This means that if we want to interpret test scores , we have to justify the validity of the interpretation , i.e. , we have to give evidence to justify that interpretation .

Gronlund (1977 : 136) states that " *We are interested in construct validity when we want to use an individual's test performance as a basis for inferring his possession of certain psychological qualities* ".

He also adds that construct validation should include the following steps :

- 1- Identifying the constructs that might account for test performance .
- 2- Formulating testable hypotheses from the theory that surrounds each construct and ,
- 3- Gathering data to test these hypotheses .

## 2- Practicality :

Oller (1979 : 16) indicates that practicality is the relationship between the resources needs and the resources available for the test . This means that the test is not practical if it requires more

resources than that available . A test is practical when it saves money , time and effort on both the tester and the testee .

### 3- Impact :

A good test should have an effect on society , educational system and the individuals . Bachman and Palmer (1996 : 29) indicate that the impact operates at two levels :

a – Micro – level in terms of the individual , and

b– Macro – level in terms of the educational system and society .

### Washback effect :

This type of effect refers to the effect that it has on the teaching that takes place before the test . Hughes (1989:1) defines washback as " *the effect of testing on teaching and learning* ". This effect may be beneficial or harmful on teaching and learning which are carried out by individuals , educational system , and society .

### 4- Authenticity :

Bachman and Palmer (1996:23) define authenticity as " *the degree of correspondence of the characteristics of a given language test task to the features of target language use task* ". This means that we have to be able to show clearly that the performance on language tests corresponds to language use in certain fields other than the language test itself . They also add that authenticity provides a means for investigation the extent to which score

interpretations generalize beyond performance on the test to language use in the target language use domain .

Authenticity is considered as an important quality of a test because of its effect on testees' perceptions of the test and on their performance .

5- Discrimination :

This feature is one of the most important qualities of a good test . It refers to the capacity of a test to discriminate among the different individuals and to see whether there are any differences in the performance of the individuals within the group . The teacher needs this feature in order to see whether his students have mastered the materials they have studied or not . In addition , he / she needs a test to discriminate so that he / she can assess the abilities and find areas of difficulty . (Heaton , 1975 : 157-8).

## Chapter Three

### Review of Previous Studies

#### 3.1 An Introductory Note

As far as assessment is concerned , a lot of work has been done in this field . It is conducted in terms of evaluating the testees' achievement . The purpose behind reviewing the studies is to found a basis for the present study and shed some light on the nature of the problem under study in order to achieve the intended aim . The related studies are presented chronologically .

##### 3.1.1 Goodson (1975)

The purpose of this study is to examine ways of using objective – referenced diagnostic tests and remedial instruction in the context of a college help session . It also attempts to assess student learning during various kinds of help session following the use of objective – referenced diagnostic tests . In addition , the study tries to investigate the influence of diagnosis and remediation on the retention of pupils .

In order to fulfil the above – mentioned goals , (79) undergraduate physical science students are chosen to one of four treatment groups and a control group .

Students are assigned to laboratory sections at the time of course registration . The sections chosen to be used in this study received the following treatments :

1. A list of objectives for each exercise studied , a diagnostic test based on these objectives , and a help session designed for reteaching incorrect responses on the diagnostic test .
2. A list of objectives for each exercise studied , a diagnostic test based on these objectives , and a help session designed to encourage students to ask questions concerning items on the diagnostic test .
3. A list of objectives for each exercise studied , a diagnostic test based on these objectives and a help session which advised students to use various resources to find answers to incorrect responses on the diagnostic test .
4. Only a diagnostic test and the help session advising students to use various resources to find answers to incorrect responses on the diagnostic test .

The control group do not receive any of the treatments mentioned above . The investigator of the study concludes that :

1. Students' achievement is greater when objective – based diagnostic tests are used to guide help session in which students are encouraged to ask questions .
2. Objective – based diagnostic testing with remediation provides greater student achievement than the lack of remediation .
3. Retention is greater when students are advised to discover on their own correct responses to objective – based diagnostic tests.

### 3.1.2 Margot (1982)

Margot studied the effect of formative assessment and corrective learning on achievement . Four experimental treatment groups of ninth and tenth grade students were used in the study . All treatment groups were given four components out of the six components of the mastery learning model (specific objectives , pretest , group instruction , post test) . One treatment group was given the formative assessment and corrective components , the second was given the formative assessment components , the third corrective , and the fourth neither formative assessment nor correctives .

The findings of the study contribute to the body of evidence that in mastery learning the feedback , corrective process increases achievement , in this case , among ninth and tenth grades low achievers .

### 3.1.3 Knight (1983)

The researcher studies the effect of diagnostic achievement testing on subsequent achievement . The study focuses on the development of diagnostic tests at a source external to the classroom for the purpose of determining whether tests developed using the best advice available from previous work would be acceptable to teachers .

Three groups , collectively representing (30) schools and (120) teachers , are chosen randomly from the population of schools with junior classes (grades 4 to 6) in an Eastern Ontario Country

jurisdiction . The Tests and Diagnostic Summary Group are given class sets of diagnostic tests based on the curriculum for the county. Teachers are requested to employ the tests at the appropriate time in instruction , and after initial scoring , to submit the tests to a central location for computer summarization . In approximately two days , a printout is returned to the school which contains summary achievement statistics , but also a visual summary of performance by skill .

The Test only Group receives the class sets of tests but no summarization . The Control Group receives no specific treatment but is not constrained in any way from normal practice .

At the end of the study period , a questionnaire to survey teacher interest and satisfaction is administered . The final instrument involves a post – test administered to all students in all groups . Results indicate that :

1. Diagnostic tests created for the study are found by teachers to be highly useful .
2. The possibility of teacher threat does not materialize as teachers contribute tests willingly .
3. Achievement of the Tests and Summary Group as measured by the post test is found to be statistically significantly higher than that of the other groups for 2 of the 3 grades .

#### 3.1.4 Thomas (1983)

The purpose of this study is to determine the effect of formative testing , diagnostic and prescriptive remediation , and

retesting on students performance in calculus in a conventional classroom instructional setting . The researcher conducts an experiment at Montana State University involving six sections .

The students are divided randomly into two groups , an experimental group and a control group . Both groups take the same quizzes and examination . The control group receives conventional forms of written feedback on their quizzes and remedial assistance based on their own perceptions of need and initiative in seeking help . The experimental group receives a detailed error analysis for each formative quiz and remediation individualized to address the diagnosed errors .

At the end of the quarter , students performance is evaluated on the sum of the three one hour examination scores plus the final examination score .

The researcher concludes that students performance in calculus can be significantly improved by using formative testing , prescribed remediation , and retesting .

### 3.1.5 Muheidat (1990)

The purpose of this study is to determine the effect of formative testing and prescriptive remediation upon student achievement in math at Kufur Asad Preparatory School for boys in Jordan involving three sections of 8<sup>th</sup> grade students .

The three sections are divided into three groups , two experimental , and one control group . The control group receives conventional forms , the first experimental group takes formative



tests only , and the second experimental group takes formative tests with remedial teaching .

At the end of the study period , students achievement is evaluated as the sum of the final achievement test scores.

The researcher concludes that students achievement can be significantly improved by using formative testing , and prescribed remediation procedure . Adding prescriptive remediation to formative testing makes only a little marginal increment in student achievement score .

### 3.1.6 Nader (1999)

The researcher investigates the attitudes of the English language teachers toward using formative evaluation in improving English language learning . He also tries to find the differences in teachers' attitudes with regard to four independent variables (teachers' gender , qualification , experience and school location) .

The researcher administers a questionnaire to all the English language teachers in Tulkaram District . The population for this study is 152 English language teachers. The researcher finds that :

1. The English language teachers have positive attitudes concerning the three domains of formative evaluation (teacher's instructional method , student's performance , curriculum design).
2. The female teachers have more positive attitudes than the males.

3. Teachers who hold less than BA have more positive attitudes than those who hold BA degree or more .
4. Teachers who have more than 10 years of experience and the teachers who have less than 5 years of experience have extremely similar attitudes but with a slight difference in favour of the teachers who have more than 10 years of experience .
5. The teachers whose schools are located in villages have more positive attitudes than the teachers whose schools are located in cities .

### 3.1.7 Al-Sa'doon (2000)

The purpose of this study is to investigate the effect of formative evaluation strategy on students' achievement in Islamic education in the second intermediate grade in Baghdad . The researcher chooses randomly two groups of students to be the sample of the study , 21 students for the experimental group and 21 students for the control group .

At the end of the study period , students' achievement is evaluated and she finds that student achievement in Islamic education is significantly improved by using formative evaluation strategy .

## 3.2 Discussion of the Previous Studies

All the previous experimental studies aimed at determining the effect of formative assessment strategy on pupils' achievement in different fields of knowledge .

These studies differ , in one way or another , in the procedures or approaches that are adopted to achieve the intended objectives . Some of the studies dealt with other independent variables , such as , remedial instruction in Goodson's 1975 , corrective learning in Margot's 1982 , diagnostic and prescriptive remediation and retesting in Thomas's 1983 , and Muheidat's , 1990. Nader's 1999 , on the other hand , investigates the attitudes of the English language teachers toward using formative evaluation in improving learning .

The samples of subjects were drawn from different populations . They were undergraduate students as in Goodson's ; ninth and tenth grade students as in Margot's . Junior classes as in Knight's ; college students as in Thomas's ; preparatory pupils as in Muheidat's ; intermediate school pupils as in Al-Sa'doon .

The number of subjects were 79 in Goodson's , 30 schools and 120 teachers in Knight's , 152 pupils in Nader's, 42 pupils in Al-Sa'doon's .

The sample of the present study is 70 male pupils . They are chosen from the second year intermediate school pupils , i.e. , pupils who have had three years of English .

The number of treatment groups was not the same for all the studies . Five treatment groups were involved in Goodson's study , (four experimental and one control) . Margot's study involved four experimental treatment groups. Knight's used two experimental groups and one control . Thomas's used two groups ; an experimental group and a control group . Muhaeidat's study

involved three groups ; two experimental and one control . Al-Sa'doon's study used two treatment groups ; one experimental and one control .

In comparison , two groups are used in this study : one is experimental and the other is control .

The technique used by most of the studies to ensure equivalence of treatment groups was simple randomization. This technique was used in the studies of Knight , Thomas and Al-Sa'doon . The other studies did not refer to such a matter .

In the present study , equivalence of the experimental and control groups is achieved by choosing two intact classes randomly , and in the following variables : scores in English in the previous academic year , scores in English in the first month of this year , age and parents' level of education . It is clear that group equivalence is more seriously dealt with in the present study than in any study of those mentioned in this chapter .

The kind of experimental design was not mentioned in the pervious studies . The present study utilizes the (post-test only group design) .

The instruments used to measure the dependent variable , i.e., achievement in the most previous studies were written achievement tests . The present study also utilizes a written achievement test to measure the pupils' achievement . The total number of items is forty-eight items . The achievement test consists of five questions in the light of the content analysis . Four sections have been evaluated throughout the achievement test . These

sections are : structure , vocabulary , reading comprehension and composition .

The present study differs from the previous studies in that it compares between the scores of each section of the experimental group and control group . Then , it compares between the entire achievement of both groups .

In only one of the surveyed studies (Al-Sa'doon's) the researcher herself taught the treatment groups . The other studies did not tell who did the teaching .

In the present study , the researcher himself taught the subjects of the two groups .

The surveyed studies used different statistical tools to test the hypotheses . Analysis of covariance was used in Muheidat's . t-test is used in Al-Sa'doon's .

The present study uses t-test as the main statistical tool . None of the previous studies investigates the effect of formative assessment strategy in the field of teaching English as a second or a foreign language . This is the main reason that attracts the attention of the researcher to select his topic .

## Chapter Four

### Procedures

#### 4.1 An Introductory Note

The researcher has conducted an experiment to fulfil the aims of the present study . Christenson (1980 : 35) defines an experiment as “ *Objective observation of phenomena which are made to occur in controlled situation in which one or more factors are varied and the others are kept constant* ” .

This chapter includes the procedures and the detailed steps followed by the researcher to achieve the objectives of the study .

These procedures are :

1. The experimental design .
2. Sample selection .
3. Controlling extraneous factors .
4. The instruction material .
5. instruction .
6. The research instruction including :
  - The achievement test .
7. The statistical instruments used .

#### 4.2 The experimental Design

The experimental design has a great value since it helps the researcher to follow the right way to get constant results that may assist the researcher to solve the problem of the study and to verify his hypotheses . (Kerlinger 1978 : 275) , and to obtain valid and

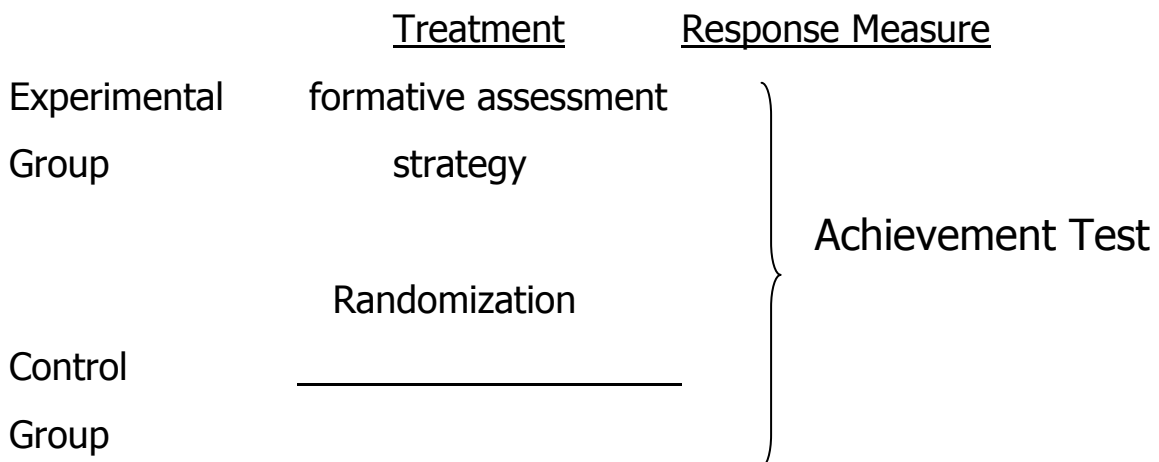
suitable answers to the study questions (Christensen , 1980 : 158) and (الزوبعي و الغنام ١٩٨١ : ٩٤- و ما بعدها) .

As far as educational experimental design is concerned, it is essential for any researcher to control the whole variables as accurately as possible . (Vandaline , 1985 : 381)

Since the aim of the present study is to assess the effect of the independent variable on the pupil's achievement , the researcher has chosen the " *post – test only control group design* " which is widely used in the educational research . (Robinson , 1981 : 228 ; Popham , 1975 : 210 ; Campbell and Stanley , 1963 : 25 ) .

The researcher has used this design on the basis that his design assumes that the two groups ; the experimental and the control groups are correlated groups in which two groups of pupils are randomly assigned , and to avoid the bad effects that may affect the results in case of using the pre-test on an instruction course lasted for only two months. (Mouly , 1978 : 253) .

This design can be drawn as follows :



(Christensen , 1980 : 170-1)

In this design , only the experimental group receives the independent variable , i.e. formative assessment strategy . Then the scores of both groups are compared . If the experimental group's scores are significantly higher than the scores of the control group , the difference belongs to the independent variable (Robinson, 1976 :189).

The experimental group is exposed to the formative assessment strategy , whereas the control group is not exposed to such strategy .

### 4.3 Sample Selection Procedures

The population of the present study is the second year intermediate school pupils in day time intermediate and secondary schools for boys in Baghdad in the academic year 2002-2003 .

There are four general directorates of Education in Baghdad . They are :

1. The General Directorate of Education in Al-Rusafa , the First .
2. The General Directorate of Education in Al-Rusafa , the Second .
3. The General Directorate of Education in Al-Karkh , the First .
4. The General Directorate of Education in Al-Karkh , the Second.

The researcher has chosen the General Directorate of Education in Al-Rusafa , the First , which contains 20019 second grades . This directorate is divided into five sectors , they are :

- 1- Center of Al-Rusafa Sector with 1363 second-graders .



- 2- Adhamiya Sector , with 2895 second-graders .
- 3- Al-Sha'ab Sector with 3746 second-graders .
- 4- Al-Rashdiya Sector with 1418 second-graders .
- 5- Al-Sader City Sector , with 10597 second-graders .

The researcher has chosen Al-Sader City Sector randomly . This sector includes (84) intermediate and secondary schools for boys . Al-Irtiqā'a Intermediate School has been randomly chosen . This school includes 156 second graders divided into four sections ; A,B,C, and D . Two sections out of the four are selected randomly to be the sample of the study . In the same way section (A) which includes 37 pupils was chosen to be the experimental group , and section (D) which contains 38 pupils to be the control group . The total number of the pupils in both sections is (75) pupils . The scores of the repeaters are excluded from the experiment , but they are kept in their classes during the period of the experiment . Two pupils of section (A) are excluded because they are repeaters and one pupil of section (D) is excluded for the same reason and two pupils left the school to join an evening school . Hence the final number of the sample subjects is (70) pupils (35) in each group . (See Table 1)

Table (1)

The distribution of the subjects of the study

Group	Section	Pupils No.	Repeaters	Final No.
Experimental	A	37	2	35
Control	D	38	3	35
		75	5	70

The researcher is fully convinced that the pupils of the two groups are extraordinarily similar since they study in the same school and they are from the same district , moreover; they are distributed among sections randomly not on their attainment level or any other factors .

Nevertheless , the researcher has tried to control some variables that may affect the outcome of the experiment . As far as equalization is concerned , the researcher matches the subjects of the two groups in the following variables :

1. The level of fathers' education .
2. The level of mothers' education .
3. The pupils' scores in English in the previous year.
4. The subjects' age .
5. The pupils' scores in English in the first month of the academic year 2002-2003 .

The information mentioned above is taken from the personal school records of the pupils , the school records of the pupils' scores and from the pupils themselves.

The differences are tested at (0.05) level of significance using Chi-Square Test and t-test formula for two independent samples under (68) degree of freedom .

#### 4.3.1 Fathers' Level of Education

To find out whether there are significant differences between the two groups in the level of fathers' education , the Chi-Square test is used and the calculated value is (2.701) , while the table

value is (9.49) under d.f. (4) . So , the difference is not significant .  
(See Table 2)

Table (2)

Chi-Square and other Statistics for the level of fathers' Education

Group	No.	Illiterate	Primary	Intermediate	Secondary	Diploma	University	d.f.	Chi-value	
									Computed value	Table-value
Experimental	35	7	6	6	3	6	7	4	2.701	9.49
Control	35	2	8	6	8	4	7			
Total	70	9	14	12	11	10	14			

#### 4.3.2 Mothers' Level of Education

It is found out that there is no significant difference between the two groups as far as mothers' level of Education is concerned. The calculated value is (0.437) while the critical value is (7.81) under 3 d.f. at (0.05) level of significance . Hence the difference between the two groups is not significant . (See Table 3)

Table (3)

Chi-Square and other Statistics for the level of mothers' Education

Group	No.	Illiterate	Primary	Intermediate	Secondary	Diploma	University	d.f.	Chi-value	
									Computed value	Table-value
Experimental	35	5	10	7	2	3	5	3	0.437	7.81
Control	35	4	11	5	4	5	6			
Total	70	9	21	12	9	8	11			

#### 4.3.3 The Age of Subjects (measured in months)

The differences among the pupils of the two groups are not significant as far as this variable (the age of the subjects) is concerned because the calculated t-value is (0.456) whereas the critical t-value is (2.000) . (See Table 4)

Table (4)

t-value and other Statistics for the Age of the Subjects

Group	N	$\bar{X}$	S <sup>2</sup>	d.f.	t-value	
					CTV	TTV
Experimental	35	162.8	55.817	68	0.456	2.000
Control	35	162.057	37.254			

#### 4.3.4 The Subjects' Level of Achievement in English in the first intermediate Grade

The mean scores of the experimental group and the control group in the first year intermediate grade are (57.743) and (58.714) respectively . The t-test revealed that there is no significant difference between the two groups at (0.05) level . The calculated t-value is (0.504) , whereas the tabulated value is (2.000) which indicates no significant differences . (See Table 5)

Table (5)

t-value and other Statistics for the pupils' level of Achievement in English in  
First Intermediate Grade

Group	<i>N</i>	$\bar{X}$	$S^2$	d.f.	t-value	
Experimental	35	57.743	62.762	68	CTV	TTV
Control	35	58.714	67.233		0.504	2.000

#### 4.3.5 Pupils' scores in English in the first month of the Second Intermediate Grade

After applying the t-test , the experimental group and the control group are equivalent in the pupils' scores of the first month in English . The calculated t-value is (0.364), while the tabulated t-value is (2.000) which is not significant . (See Table 6)

Table (6)

t-value and other Statistics for the pupils' level of Achievement in English in the first month of the Second Intermediate Grade

Group	$N$	$\bar{X}$	$S^2$	d.f.	t-value	
Experimental	35	53.429	157.045	68	CTV	TTV
Control	35	52.429	107.788		0.364	2.000

#### 4.3.6 The Extraneous Factors Jeopardizing External and Internal Validity

To avoid confusing the outcomes of the study , the researcher has tried to manipulate the influence of extraneous factors that may affect the dependent variables, other than the independent one (Christensen 1980 : 94) . Hence the researcher has determined the independent variable (formative strategy) and the dependent variable (the pupils' achievement) , to control the extraneous variables so as to get favourable internal validity .

The most prominent variables that may have direct influence on the experiment are :

##### 4.3.6.1 History

This factor refers to certain events that may happen during the experiment and have their effects on the final result of the experiment . (Lewin , 1979 : 338) . Nothing happened during the period of the experiment that may upset it , so this factor has been restrained .

#### 4.3.6.2 Maturation

It refers to the internal changes that occur in the individual conditions during the passage of time , whether they are biological or psychological growth , such as age , learning , boredom , hunger and the like . (Cambell and Stanley , 1963 : 5).

This factor was controlled since the period of the experiment lasted for only two months , i.e. the period was not too long to attribute the influences to maturity during the experiment . In addition , if there is an effect on the results , the two groups will be affected equally .

#### 4.3.6.3 Selection Bias

Selection bias is clearly seen when a differential selection procedure is used for placing individuals in the various comparison groups . (Christenson 1980 : 98)

The researcher followed a random selection of the sample of the present study . So two classes were selected randomly . Moreover , an appropriate statistical procedure has been used to equalize the two groups in some variables to avoid any influence of any external or internal factor .

#### 4.3.6.4 Experimental Mortality

Lewin states that the experimental mortality is the absence of subjects during the period of the experiment for different reasons , such as death , accidents and the like (1979 : 60) .

The subjects of both groups did not encounter the effect of such factor during the two months of the experiment since all the subjects participated in the experiment except some pupils in both groups who were absent for some lessons which is considered natural in such a state in the educational field .

#### 4.4 Instrumentation

One of the most important factors on which the success of the experiment depends is the instruments that can be used to measure the dependent variable . (Christensen , 1980 : 97) . The researcher has measured the pupils' achievement by constructing a printed test as well as using the same scoring plan for both groups . The researcher does not tell the subjects that they are in an experiment in order –for them- not to be affected in any way by such knowledge .

#### 4.5 The Instructional Materials

Eight Units from the English text-book for the second intermediate grade from NECI : Book 4 are chosen . These eight units are : units 6 to 13 . Units from 1-5 are excluded because they are review units of Book 3 . The pronunciation skills have been exempted in the achievement test for reasons that have to do with the practicality of examination and the availability of time . In addition to that , there are no English language laboratories available in most intermediate schools that will make the results of experimentation and examination in this language skill reliable .



Consequently , the block of the achievement test items depends heavily on the oral practice because the time allotted to it is (30) minutes , i.e. about (75%) of the time of each lesson is devoted to oral practice (Al-Hamash et al. 1990 : 6) and (15) minutes should be devoted to the other activities (Al-Hamash 1984 : 122) .

The content of the oral practice exercises is limited to:

1. past perfect tense with 'before' and 'after' .
2. present perfect continuous tense with " since " and ' for ' .
3. passive voice " singular and plural " in the positive and negative .
4. if-clause (type two and three) .
5. indirect speech .
6. contact relative clauses .
7. please + verb + (pro.) noun + to + infinitive .
8. the relative pronouns (which , that , whose) .
9. out of .
10. during .
11. enough .
12. would you mind + gerund .
13. would like and wouldn't like .
14. unless .
15. therefore .
16. half .
17. third .
18. prepositional phrases with (on , in) .
19. -ing participle phrase modifying the noun .

20. what for .
21. I'm afraid .
22. 'could' as a polite request .
23. except .
24. still .
25. however .
26. ` may ' for asking and giving permission .
27. what about + gerund .
28. must + passive voice .
29. present perfect tense in the passive form .

#### 4.6 The Experiment Application

The instruction of both groups , the experimental and the control groups , started on 2<sup>nd</sup> of October and ended on the 30<sup>th</sup> of December of the academic year 2002-2003 .

The researcher himself taught the two groups to avoid the teacher variable . It is noteworthy that the researcher has an experience of teaching English for about 20 years in the intermediate and secondary schools . The arrangement of the lesson for the two groups was on Saturday , Sunday , Monday , Wednesday and Thursday every week .

Both groups were taught according to the approach followed in Iraqi intermediate schools (audio-lingual approach) . Formative assessment strategy was followed in testing the experimental group (formative assessment group) , i.e. , the experimental group was tested informally and continuously at the end of each unit , whereas

the control group was tested at the end of the experiment . The researcher plans each lesson after discussing it with the teachers of English in the school where experiment is conducted (See Appendix A) . The distribution of the classtime in the plan was as follows :

Table (7)  
Time Allocation to the Various Activities

Activity	time allotted in minutes
Dialogue	4
Oral Practice	15
Oral Practice	15
Pronunciation	3
Reading	4
Written Homework	4
	45

#### 4.6.1 The Control Group

The researcher taught this group using the audio-lingual approach steps as recommended in the Teacher's Guide .

These steps are :

##### 1. Recognition :

In this step, the teacher introduces the new material while the learners are listening with books closed. Then , he explains the new vocabulary items , difficult words are explained as well .

## 2. Imitation :

To imitate what has been recognized on the part of the pupils, the researcher reads the sentence and asks the pupils to repeat after him chorally , in groups and then individually .

## 3. Production :

This is conducted to encourage the pupils to produce sentences similar to those that have been drilled helping the learners by giving them clues .

### 4.6.2 The Experimental Group :

The procedures applied to the Experimental Group are the same as the procedures followed in teaching the Control Group except for assessing the pupils' achievement at the end of each unit by using the formative strategy .

The investigator prepares (8) quizzes to evaluate the learners' achievement of the Experimental Group . A period of about (8-10) minutes is offered to the pupils to take each quiz related to the inhand unit .

After the application of each quiz , the papers are corrected and marked . Notes are written on the papers to inform the pupils of the mistakes they have made . The idea behinds this procedure is to provide the pupil with feedback to the errors he has made , the correct answers to the questions he has failed to answer correctly . These problematic areas are noted by the teacher to be dealt with in the subsequent lessons if they are detected on a large scale .

Otherwise , i.e. if they are individual the subject is advised as to how to deal with them .

#### 4.6.3 Instruments of the Research

If one analyzes the textbooks of the intermediate grade of NECI , he will not find specific objectives for each level , i.e., they are adopted for the intermediate stage as a whole . This means that one cannot isolate the objectives of the first intermediate level from those of the second or the third intermediate levels . In other words, the objectives assigned by the Ministry of Education are not distinct (Al-Hamash , 1984 : 35-f) to be followed in expressing clearly the specific objectives (Appendix B) . Nevertheless , the researcher has taken these general objectives into account , depending heavily on the content of the textbook available in the Teacher's Guide in formulating the achievement test .

#### 4.6.4 The Achievement Test

Any instructional activity that the teacher makes inside the classroom requires different types of evaluation to follow exactly the pupils' progress and the extent to which his teaching has been effective (Thorndike , 1969 : 116) . Besides , achievement tests are the most useful means by which the teacher can diagnose the problematic areas , help the pupils to know exactly their level and to increase motivation . (Callahan , 1983 : 220 ; Harris , 1969 : 3 ; Mager,1967 : 13) .

The researcher has constructed a suitable achievement test . For the purpose of the research , it has been found necessary for the researcher to construct a teacher – made achievement test for the experiment since none of the available tests is adequate enough to realize the objectives of the research . The aim of constructing an objective achievement test is to trace the effect of formative assessment strategy on the Experimental Group subjects compared with the achievement of the Control Group subjects who are not exposed to such a strategy of assessment .

#### 4.6.5 Test Construction

Any researcher has to have a base on which he can construct a test , otherwise , the test will be weak or unserviceable . Hence , the researcher analyzed carefully the content of the eight units involved (from unit 6-13) to be the material of constructing the achievement test of the present study . The researcher analyzed the content of the oral practice of these eight units as far as structures and the vocabulary items are concerned .

Items of high frequency in the oral practice are included in the production task only , while those of low frequency are included both recognition and production tasks .

The reason behind this is that the structural approach followed in TEFL in Iraq concentrates on repetition and habit formation . Thus , it will be difficult for the pupils to remember and produce unless an item is sufficiently taught and repeated .

The achievement test also includes a passage for reading comprehension written by the researcher depending on the vocabulary items that have already been taught previously .

One of the general objectives adopted by the Ministry of Education is to enable the pupils to read simple English passages with comprehension (Al-Hamash , 1984 : 36) .

Writing is one of the most important means by which the pupil can express ideas and knowledge within limited topics and vocabulary . This involves training in writing guided composition on certain topics , one of them is writing simple paragraphs of self expression in which the pupils talk about themselves , their family , school and city or town . (Ibid : 51) . Therefore , the researcher has chosen a guided composition to be one of the questions of the achievement test to fulfil the above-mentioned objective adopted by NECI in TEFL in Iraq .

The achievement test consists of five questions and the total number of the test items is forty-eight . (Appendix C)

#### 4.7 The Written Test Validity

One of the most important features of a valuable test is validity . It is also an essential factor that should be taken into account in constructing a test . Validity refers to the extent to which test results serve their intended use (Gronlund , 1976 : 130) . Generally speaking , validity is related to how well the test does what it is supposed to do , i.e., the accuracy with which a set of test

scores measures what it is supposed to measure (Valette , 1967 : 30 ; Heaton , 1975 : 135 ; Oller , 1979 : 4) .

Most testers have used different terms to identify validity . Some of these testers use three main types of validity ; rational , empirical and construct validity , others use internal and external validity (Alderson , et al. , 1995 : 171) . Internal validity relates to studies of the perceived content of the test and its perceived effect , while external validity relates to the studies comparing students' scores with measures of their ability gathered from outside the test .

Internal validity includes face validity , content validity and response validity . External validity includes concurrent validity and predictive validity . The third type of validity is the construct validity which is considered by some testers to be an umbrella term for internal and external validity . (Ibid :72)

The most essential types of validity in relation to constructing an achievement test are : content validity and face validity (Harris , 1969 : 21) . Hence , content validity and face validity are considered in the construction of the test for this study .

Content validity is a matter of determining whether the sample is representative of the larger universe it is supposed to represent . In other words , content validity is how well the test measures . According to Oller " *content validity is related to the question of whether the test requires the examinee to perform tasks that are really the same as or fundamentally similar to the sorts of tasks one normally performed in exhibiting the skill or ability that the test purports to measure* " (Oller , 1979 : 50) . The achievement test has



content validity , since it is constructed carefully depending on the analysis of the skills the course includes . " *The test has content validity if it is designed to measure the mastery of a specific skill or the content of a particular course of the study*". (Harris , 1969:18) .

Face validity refers to the acceptance of the apparent ability of a test or study to measure what it appears to measure (Page et al.,:131) , i.e., face validity involves intuitive judgments rather than statistical methods .

To consolidate face validity of the achievement test of the present study , the test items are submitted to a jury of experts . \* (See Appendix C)

The jury members have discussed the items of the test with the researcher . They agree that the test items are appropriate to fulfil the purposes they are designed for, except for some modification and comments which are taken into consideration in the final version of the test .

---

\* The jury members are :

1. Abdulah Al-Musawi , Ph.D. Prof.
2. Abdul-Jaleel Jassim , M.A. Asst. Prof.
3. Ahmed Mustafa , Ph.D. Lecturer
4. Ayif Habeeb , M.A. Prof.
5. Dhuha Atallah , Ph.D. Lecturer
6. Fatin Khairy , Ph.D. Lecturer
7. Firas A. Ma'rouf , M.A. Asst. Prof.
8. Istiqlal Hassan , Ph.D. Asst. Prof.
9. Nahida Taha , Ph.D. Asst. Prof.
10. Sabah Al-Ajly , Ph.D. Prof.

#### 4.8 The Pilot Study

The Pilot study is administered in order to identify the time needed for the final administration of the test , determine the difficulty level and discrimination power of the test items , have an idea about the arrangement and requirements for the test , make any necessary modifications on the test , and have a clear-cut picture of the scoring scheme .

With the validity of the test ensured , it is now necessary to try out the test on a sample similar to the sample of this study chosen randomly for the purposes of the Pilot Study . The test is given to (100) pupils ; (70) pupils (section B and C) from Al-Irtiqaa' School and (30) pupils from Al-Shuhadaa' Intermediate School in the General Directorate of Education in Al-Rusafa , The First .

The Pilot Study has also provided us with valuable information regarding the applicability of the test , the efficiency of the scoring scheme and requirements for the study . The Pilot Study indicates that the time needed for the test is about an hour .

#### 4.9 The Test Reliability

After assurance of test validity , the next task to be considered in the preparation of achievement test is its reliability . " *Reliability of a test is a matter of how consistently it produces similar results on different occasions under similar circumstances* ". (Oller , 1979 : 4) , i.e., test reliability concerns the extent to which test results are repeatable . The relationship between validity and reliability is that a test cannot be valid unless it is reliable . Reliability provides the

consistency that makes validity possible , and indicates the confidence to be placed in test interpretations . (Gronlund , 1977 : 130)

Actually there are four basic methods to determine test reliability ; the split-halves method , Kuder-Richardson method , equivalent-form method and Test-Retest method (Ahman and Clock, 1975 : 237f) . Two of them are used in this research . The first is the Test-Retest method .

The Test-Retest method is used in this study . This method refers to the stability of test scores over some given periods of time. To achieve this aim , a class of thirty-four pupils chosen randomly is retested after a fortnight . Pearson Correlation Formula is used for the two sets of scores . The correlation coefficient is found to be (0.741) which is considered a high positive correlation (Hedges, 1960 : 22) . (See Appendix E)

The second method used to estimate the test reliability is the Kuder-Richardson formula with its variation 21 , which splits the test into as many as parts as there are items in the test . This method depends on the mean and variance of scores . It measures the internal consistency of the test items . It was found that the reliability coefficient was 0.923 . Tuckman (1975 : 257) believes that *" teacher-built tests are usually considered adequate with reliabilities of 0.60 or above "*.

To judge test reliability , two scorers\* are requested to score pupils' responses to see how similar results that can be obtained from assessment of the same responses are .

The marks assigned by the two teachers are not only almost exactly the same , they are also almost exactly similar to those of the researcher . Thus , the reliability of the test is also ensured .

#### 4.10 Test Administration

After assuring that the test requirements have been carried out , the test is applied to (70) pupils . Two halls are prepared identically for this purpose to avoid any impact that might happen to affect the performance of the testees . The researcher does his best to make the test as unified as possible taking into account the conditions under which the test is administrated .

The pupils do not need any additional instructions , for the test instructions are quite clear and understandable , moreover , the researcher explains the instructions in Arabic and in English .

On January 3<sup>rd</sup> , 2003 , the test was applied in two classrooms in Al-Irtiqaa' School by assistance of two teachers of English in this school . The time devoted for the test was (80) minutes .

---

\* The scorers are Ibraheem Issa and Saba Jasim Teachers of English in Al-Irtiqaa' Intermediate School .

#### 4.10.1 The Scoring Scheme

Since the distribution of marks among the items should be clear and precise to ensure the objectivity of the test , the researcher followed a scheme that ensures accurate results on which to decide the level of each individual . Questions (1 and 5) are multiple-choice questions , hence one mark is given to the right choice and zero to the wrong response . Zero is also given if the letter of the choice is not clear or the item is left unanswered by the testee . In question (2), every item is given two marks if the answer is completely correct and one mark is given if there is a structural error or morphological error . Zero is given to the item that has more than one error . Spelling and punctuation errors are ignored . Question (3) contains (4) items , item (1) and item (3) demand two correct forms , so each form is given one mark for the correct one and zero for the incorrect one. Thus , the total mark for this question is (6). Each item in question (4) is given two marks if the answer is completely right ; one mark is given for the correct order of the sentence and one mark for the correct tense , so the highest mark for the test is (60) .

The researcher himself corrects the answering sheets of the two groups ; experimental and control groups . The scores of the individuals are shown in Appendix (F) .

#### 4.10.2 Item Analysis

As stated earlier one of the aims of the pilot study is to check the difficulty level of each item in the five questions .

Madsen (1983 : 182) states that " *A test is considered too easy if more than (90) percent get it right . An item is considered too difficult if fewer than (30) percent get it right* " , Bloom et al., on the other hand , state that a test is considered good if its items difficulty level varies in difficulty from (20) to (80) (Bloom et, al., 1981 : 95 ; See also Downi, 1967 : 214f) . To find out the difficulty level for each item, the following formula is used :

$$\frac{\text{High correct} + \text{Low correct}}{\text{Total number in sample}} \quad (\text{Madsen 1983 : 181})$$

The difficulty level between 20-80 percent is adopted in this research since most specialists in this field agree on this percentage. It is advisable that the test should include easy items at the beginning of the test to encourage and motivate testees to answer . The test should also contain difficult items to help good pupils show their abilities .

The pupils' scores ranged from the highest to the lowest score, then divided into two halves ; an upper half and a lower half. This procedure is the best way that can be used in item analysis (Remmer et al., 1965 : 269) . The pupils in the upper and lower groups who chose each possible response for each test item are listed . The difficulty level of each item is calculated by the application of the item difficulty formula mentioned earlier . As a

result it is found that it ranges between (16.67) and (70.37) percentage .

Item discrimination power is the other essential procedure followed by the researcher . Ebel defines item discrimination power as “ *the difference between good and poor students in proportions of correct response* ” (Ebel , 1972 : 376 ; See also Wingard , 1981 : 184) .

The researcher finds that the item discrimination power ranges between (0) and (0.741) (See table 8) . Ebel (1972 : 406) indicates that “ *good test items should have indices of discrimination of 0.30 or more* ” .

Table (8) shows that there are weak items whether in their difficulty level or in their discrimination power .

According to Madsen (1983 : 185) :

*“ Weak distractors often cause test questions to have poor discrimination or an undesirable level of difficulty . No set percentage of responses has been agreed upon , but examiners usually feel uneasy about a distractor that isn't chosen by at least one or two examinees in a sample of (20) to (30) test papers . ”*

On the other hand , a good distractor is one that attracts more pupils from the lower group than the upper group .

(Gronlund , 1976 :268) .

It is found here that there are no weak distractors in Question (1) (the multiple-choice question) except the items whose discrimination power or difficulty level are low. Some of these items are modified and tried out again by a number of the same sample of testees. The researcher , then reanalyzes these items . The other low items are not modified because their difficulty level is a result of the difficult material itself that the pupils could not master for one reason or another . These items should be included in the test since they measure some relevant aspects of the achievement .

(Ebel, 1972 : 395)

Table (8)

The difficulty level and the discrimination power of the items

Questions	Item No.	Difficulty level %	Discrimination power
1	1	62.69	0.370
	2	61.11	0.482
	3	62.96	0.333
	4	48.15	0.556
	5	37.04	0.444
	6	55.56	0.519
	7	38.89	0.556
	8	27.78	0.333
	9	40.79	0.444
	10	44.44	0.444
	11	61.11	0.556
	12	70.37	0.333
	13	50.50	0.556
	14	42.59	0.333
	15	50.00	0.407
	16	42. 59	0.222(0.333) *
	17	25.93	0.333
	18	37.04	0.519
	19	53.7	0.333
	20	44.44	0.444
	21	55.56	0.370
	22	42.59	0.333



	23	51.85	0.519
	24	38.89	0.482
	25	40.74	0.00(0.444) *
	26	57.41	0.482
	27	53.7	0.444
	28	48.15	0.370
	29	64.82	0.704
2	30	12.96 **	0.185(0.40) *
	31	20.37 **	0.259 **
	32	18.52 **	0.222 **
	33	14.82 **	0.296
	34	16.67 **	0.333
3	35 A	16.67 **	0.333
	B	59.26	0.149 **
	36	51.85	0.519
	37 A	62.96	0.444
	B	16.67 **	0.333
	38	20.37	0.333
4	39	42.59	0.333
	40	44.44	0.444
	41	27.78	0.333
	42	48.15	0.556
	43	40.74	0.444
5	44	50.00	0.407
	45	42.59	0.333
	46	40.74	0.444
	47	59.26	0.741
	48	57.41	0.482

\* This mark refers to the low indices of difficulty level of discrimination power which resist the item modification due to the nature of the material itself , since such item is relevant to important aspects of the achievement measured by the test .

### 4.10.3 Statistical Tools

The following statistical tools are used in the present study :

- 1- t-test , to find out the significance of differences between the two groups in non-experimental variables and in the experimental variable .
- 2- Person Correlation Coefficient , to estimate the reliability of the achievement test using the test-retest method .
- 3- Chi-square to find out the significance of differences between the two groups as far as the levels of parents educational background are concerned .

## Chapter Five

### Results

#### 5.1 An Introductory Note

In order to realize the aim of the present study and verify its null hypothesis which reads as follows : (There is no statistically significant difference between the mean score of the achievement of the Formative Assessment Group and that of the Summative Assessment Group in the post test) . The data obtained by administering the achievement test to the subjects of the study is statistically analyzed . The achievement test is divided into four sections : structure , vocabulary , composition and reading comprehension to see whether the differences between the experimental and control groups in these four sections are significant or not , and then the difference between these two groups in the Achievement Test as a whole .

#### 5.2 Achievement in Specific Components of the Test

##### 5.2.1 Achievement in Structure :

In order to find out whether the difference in the subjects' performance in structure is statistically significant or not , the t-test is used . The results are displayed in Table (9) .

Table (9)

t-test statistics for the study groups' achievement in structure

Groups	N	$\bar{X}$	S	d.f.	t. Value		Level of significance
Experimental	35	19.971	6.692	68	CTV	TTV	0.001
Control	35	13.428	5.812		4.367	3.460	

As the table shows , the mean score of the experimental group is 19.971 , whereas the mean score of the control group is 13.428 . Using t-test formula for two independent samples , it has been found that the computed t-value is 4.367 which is larger than the tabulated t-value , namely , 3.460 at 0.001 level of significance under 68 degrees of freedom . This means that the difference between the two mean scores is significant in favour of the experimental group as far as structure is concerned .

This shows that the improvement of the pupils' performance in structure is due to the formative assessment strategy used in accordance with Structural Approach .

### 5.2.2 Achievement in Vocabulary :

As Table (10) illustrates, the mean score of the control group is 3.571 and that of the experimental group is 4.714 . Applying t-test formula reveals that the calculated t-value is 3.764 and the tabulated t-value is 3.460 which is smaller than the calculated t-value at 0.001 level of significance under 68 degrees of freedom . The difference between the two mean scores is significant in favour of the experimental group .

This illustrates that the experimental group performance in vocabulary is better than that of the control group .

Table (10)

t-test statistics for the study groups' achievement in vocabulary

Groups	N	$\bar{X}$	S	d.f.	t. Value		Level of significance
					CTV	TTV	
Experimental	35	4.714	1.318	68	CTV	TTV	0.001
Control	35	3.571	1.219		3.764	3.460	

### 5.2.3 Achievement in Composition :

To see whether the difference between the mean score of the experimental group and the control group is significant or not , t-test formula for two independent samples is used . The figures in Table (11) show that the mean score of the experimental group is 6.657 whereas the mean score of the control group is 5.057 . It has been found that the calculated t-value is 4.50 which is greater than the tabulated t-value , 3.460 , at 0.001 level of significance under 68 degrees of freedom . The difference between the two mean scores is significant in favour of the experimental group .

This proves that using formative assessment strategy leads to improvement of pupils' achievement in composition.

Table (11)

t-test statistics for the study groups' achievement in composition

Groups	N	$\bar{X}$	S	d.f.	t. Value		Level of significance
Experimental	35	6.657	1.643	68	CTV	TTV	0.001
Control	35	5.057	1.661		4.050	3.460	

#### 5.2.4 Achievement in Reading Comprehension :

In order to investigate the significance of the difference between the two groups' achievement in reading comprehension , t-test formula is used . As shown in Table (12) the mean score of the experimental group is 4.228 whereas the mean score of the control group is 3.142. It has been found that the computed t-value is 4.527 which is bigger than the tabulated t-value which is 3.460 at 0.001 level of significance under 68 degrees of freedom . This means that the difference between the two means scores is significant in favour of the experimental group . It is clear enough that the improvement of pupils' achievement as far as reading comprehension is concerned is attributed to applying formative assessment strategy .

Table (12)

t-test statistics for the study groups' achievement in reading comprehension

Groups	N	$\bar{X}$	S	d.f.	t. Value		Level of significance
Experimental	35	4.228	0.877	68	CTV	TTV	0.001
Control	35	3.142	1.115		4.527	3.460	

### 5.3 The Entire Achievement Test

As shown in Table (13) , the mean score of the experimental group in the achievement test as a whole is 35.571 whereas the mean score of the control group is 25.200 . The t-test for two independent groups yields a computed t-value of 5.592 which is bigger than the tabulated t-value which is 3.460 , at 0.001 level of significance under 68 degrees of freedom . This means that the difference between the two means scores is significant in favour of the experimental group . Therefore , the null hypothesis stated above is rejected and the alternative hypothesis is accepted . Hence the aim of the present study, namely , to investigate the effect of formative assessment on the pupils' achievement in English is achieved . This affirms that using formative assessment strategy systematically has strong positive influence on the pupils' achievement .

Table (13)

t-test statistics for the study groups' achievement in entire achievement test

Groups	N	$\bar{X}$	S	d.f.	t. Value		Level of significance
Experimental	35	35.571	8.197	68	CTV	TTV	0.001
Control	35	25.200	7.295		5.592	3.460	

## 5.4 Interpretation

As shown in 5.2 , the effect of formative assessment strategy is positive since the difference between the mean scores of the two groups is significant and in favour of the experimental group . In other words , using formative assessment strategy affects the pupils' achievement positively and their learning is improved .

As proved empirically , applying formative assessment strategy contributes effectively in learning English as illustrated below :

1. Exposing pupils to formative testing helps them organize their time and effort . In other words , the learners will guess the study time which should be devoted to preparing the taught material at the end of each unit .
2. Repetition of quizzes at the end of each unit could belittle , to some extent , the anxiety of examination . Consequently , pupils behave well in formal testing . This will lead to better achievement .
3. Providing feedback to the pupils helps them realize their levels , their weaknesses which have to be improved and their strengths that have to be reinforced .
4. Exposing pupils to many quizzes during the teaching-learning process enhances their motivation to learn and continue their study , which in turn leads to increasing their achievement .
5. The higher achievement of the experimental group is attributed to the formative assessment strategy which gives a chance to the pupils to pay due attention to the materials taught



or discussed inside the classroom since they know that they will be tested at the end of the unit .

6. Correction of the pupils' errors on their test papers on the part of the teacher gives them the chance and the experience to remedy such errors when they are exposed to the same situations in future , and this will lead to self-confidence .

The result of the present study is in agreement with the results of certain related studies mentioned in Chapter Three , all of which have one scope in common , which is the effect of formative assessment strategy on the pupils' achievement . The present result agrees with Knight's (1983) in which the achievement of the formative assessment group as measured by the post test was found to be significantly higher than that of the other groups , and diagnostic tests created for the study were found by teachers to be highly useful . It is also in agreement with Thomas' (1983) in which it was revealed that students' achievement in calculus can be significantly improved by using formative testing .

The present result agrees also with Muheidat's (1990) in which students' performance was improved by using formative testing . The result of the present study accords also with Nader's (1999) in which formative assessment strategy improved the English language learning of the study subjects . It also agrees with Al-Sa'doon (2000) in which students' achievement in Islamic Education was significantly improved by using formative assessment strategy .

## Chapter Six

### Conclusions , Recommendations and Suggestions for Further Research

#### 6.1 Conclusions

In the light of the results of the present investigation and the researcher's own observations , the following conclusions can be reached with reasonable confidence :

1. Some teachers of English in Iraqi schools do not pay much attention to formative assessment strategy in teaching – learning process . Applications of formative strategy would increase students' achievement level .
2. Providing pupils with feedback facilitates their learning of English Language .
3. Applying formative strategy helps pupils lessen anxiety before examinations and helps them realize how and what to study .
4. Using formative testing strategy motivates pupils and reinforces their learning which leads to increase their achievement .
5. The application of formative strategy helps pupils know their progress through the learning task , and the pupils will be more competent than those who lack such feeling .

## 6.2 Recommendations

On the basis of the results of the present study strengthened by the corroborative findings of related previous studies , the investigator recommends the following :

1. Devoting at least ten minutes of each lesson period to re-teaching the previous material that the teacher thinks it needs to be re-taught . Recycling is a teaching technique that is highly recommended but very rarely used . Experience tells us that this strategy , even though applied in Book 8 of NECI through the check points , is never properly utilized .
2. Increasing English lessons to six instead of five in every week to give the teacher enough time to make quizzes , and reteach the problematic areas that the pupils encounter during teaching – learning process .
3. Training teachers of English in the use of formative strategy and asking them to follow the instructions related to such strategy .
4. Asking teachers of English to record the score of each quiz against each individual's name to follow his progress easily .
5. Asking teachers of English to write down notes and corrections of errors on the test papers of the individuals .
6. Concentrating also on oral feedback inside the classroom .

### 6.3 Suggestions for further Research

One important function of a research is to generate more researchable problems . In the light of the results of the present study , the researcher suggests the following topics to be studied :

- 1) A Study is needed to investigate the effect of formative strategy on achievement of pupils in the other classes of intermediate stage .
- 2) A study using the same procedures can be conducted in secondary schools .
- 3) Other studies can be carried out to investigate the effect of formative strategy on learning other activities, such as dialogue , pronunciation , spelling .
- 4) A similar study can be conducted in schools for girls .
- 5) A further research is needed to examine the effect of formative assessment strategy on the pupils' achievement in subjects other than English .
- 6) A survey study can be carried out to investigate how systematic application of formative strategy affects other important pupils' characteristics such as study habits , test anxiety , learning style and motivation .

## Abstract

Assessment plays a fundamental role in the teaching – learning process due to its importance for the success of this process . Without assessment one cannot know much about the fulfilment of the specific and the general objectives , neither can he follow the pupils’ development in their learning and responses . Teachers also make use of assessment in deciding whether their method of teaching is successful or not . Besides , parents know much about their children’s level through evaluation made inside schools . Finally , assessment helps in grading pupils from one stage to another and in knowing the individual differences .

Hence , the present study aims at investigating the effect of one kind of assessment ; namely the formative assessment , on the pupils’ achievement in English through an empirical investigation . For this purpose the following hypothesis is set up (There is no statistically significant difference between the mean score of the achievement of the Formative Assessment Group and that of the Summative Assessment Group in the post test) .

The study is limited to the male pupils of the second intermediate grade in Baghdad during the academic year 2002-2003 and the instructional material is limited to structure , vocabulary , reading comprehension and composition activities in Book 4 (NECI) .

To achieve the aim of this research , an experiment is designed involving two groups : Experimental and Control . Two sections of thirty – five pupils each were selected randomly from Al-Irtiqaa’ Intermediate School for Boys . One of them is randomly chosen to be the experimental group (evaluated by formative assessment) and the other to be the control group (evaluated by summative assessment) . The subjects of both groups are statistically matched in the following variables : The level of parents’ education , age , achievement in English in the previous year , and in the first month of this year .

The investigator himself taught the two groups the same material which is units 6 to 13 of Book 4 (NECI) . The experiment lasted for eight weeks .

The researcher constructed an achievement test of 48 items in the light of a careful content analysis of the material taught during the experiment . The test was exposed to a jury of experts to verify its face validity . Its reliability was calculated by using the test – retest method and Kuder Richardson method . The correlation coefficient is found to be 0.791 and 0.923 respectively . An item analysis was carried out to determine the difficulty level and the discrimination power of the items . Then the test was administered to both groups at the end of the experiment .

The t-test formula for two independent samples is used to test the significance of the difference between the two mean scores of the two groups . This difference was found to be significant at 0.001 level in favour of the experimental group . t-test formula was also used to test the significance of the differences between the two mean scores of the two groups in : structure , vocabulary , reading comprehension and composition. These differences were found to be significant in favour of the experimental group in these four sections .

On the basis of the study results , the researcher concludes that using formative assessment strategy improves pupils' achievement in English and reinforces their learning .

The study ends up with a number of recommendations and suggestions based on the findings of this study .

## Appendix A

### A model Lesson Plan for the Control Group

Class = Second Intermediate Grade

Date = October 16<sup>th</sup> , 2002

Material = -The Oral Practice Exercises 8.9 and 8.10 that introduce  
           " *have to* " in positive and negative statement .  
           -The third paragraph of 8.14 is also introduced .

#### **Objectives :**

A : Special objectives

- 1- Teaching the pupils what we mean by " *have to* " and " *don't have to* " .
- 2- Teaching the pupils to express the obligation or necessity , and teaching them how to express the absence of obligation .

B : Behavioural Objectives

- 1- The pupil will be able to recognize " *have to* " and " *don't have to* " .
- 2- The pupil will be able to use " *have to* " and " *don't have to* " in sentences .

8.9: " *have to* " in positive statements . (15 minutes) .

## **Techniques :**

### 1. Presentation :

The researcher says the following sentences orally . He writes them on the blackboard and explains the meaning of the new vocabulary items .

1. The blackboard is dirty . You have to clean it .
2. It's hot . He has to open the window .
3. It's late . We have to go now .

### 2. Imitation :

The pupils are asked to repeat these sentences chorally and then individually after the researcher .

### 3. Production :

The researcher conducts a substitution drill by giving cues . The pupils have to substitute the cue word into their appropriate place in the sentence .

T: Ali has to clean the blackboard .

T: Muna wash the dishes .

P: Muna has to wash the dishes .

T: help her mother .

P: Muna has to help her mother .

T: The girls help the teacher .

P: The girls have to help the teacher .

T: They sleep early .

P: They have to sleep early .

T: He

P: He has to sleep early .





**Reading :**

## 8.14 Section 3

The researcher reads this section aloud with the pupils' book open . Then, he explains the meaning of the new words or expressions . The next step is that the researcher reads the section again and asks the pupils to repeat after him . He asks the pupils to read this section silently . He asks individual pupils to read the section aloud . After that , he asks the class questions about this section . The following questions are included :

1. What did he buy the next day ?
2. Where did he decide to go at ten ?
3. How did he go to the university ?
4. What had he done before he went to the first bus?

## Appendix B

### The objectives of the programme in the Intermediate Stage

The aims of the English language programme in the Intermediate Stage are :

- 1- Using English effectively in understanding and speaking English language at normal speed within a limited vocabulary .
- 2- Reading simple English passages with comprehension and enabling the pupils to ideas and knowledge in writing within limited topics and vocabulary .

(Al-Hamash , 1984 : 35-f)

## Appendix C

### The Initial Version of the Achievement Test given to the Jury

University of Baghdad ,  
College of Education (Ibn-Rushd) ,  
Department of Educational and  
Psychological Sciences ,  
Higher Studies

**Dear Sir ,**

The researcher is conducting an experimental study entitled " *The Effect of Formative Assessment on the Pupils' Achievement in English* " .

This study is limited to the second intermediate grade . The material is restricted to oral practice , reading and writing . To achieve the aims of the study , the researcher has prepared a test of (48) items . The pupils are expected to organize and produce the content presented in the textbook after being taught eight units in Book IV of (NECI) , (from unit 6 to unit 13) , because the first five units of Book IV review Book III .

The content of oral practice exercises is enclosed with the test to save the effort of looking for the content in the textbook .

I will be grateful if you , as a specified in the field of TEFL , Linguistics, or applied linguistics , advise me as to the suitability of the test items . Your suggestions will be highly appreciated .

*Thank you in advance for your kind  
Co-operation and valuable comments .*

Yours sincerely  
Ph. D. Candidate  
Ridha Ghanim Al-Saedi

***The content of oral practice exercises is limited to :***

1. past perfect tens with "*before*" and "*after*"
2. present perfect continuous tense with "*since*" and "*for*"
3. passive voice (singular and plural) in the positive and negative
4. if-clause (type two and three)
5. indirect speech
6. contact relative clauses
7. please + verb + (pro.)noun + to + infinitive
8. the relative pronouns (which , that , whose)
9. out of
- 10- during
- 11- enough
- 12- would you mind + gerund
- 13- would like and wouldn't like
- 14- unless
- 15- therefore
- 16- half
- 17- third
- 18- prepositional phrase with (on , in)
- 19- ing participles phrase modifying the noun
- 20- what for
- 21- I'm afraid
- 22- could as a polite request
- 23- except
- 24- still
- 25- however
- 26- " may " for asking and giving permission
- 27- what about + gerund
- 28- must + passive voice
- 29- present perfect tense in the passive form .

**QI- Write the number of the sentence and the letter of most suitable choice :**

- 1- It is hot in Iraq \_\_\_\_\_ the summer .  
 a. through    b. round    c. during    d. between
- 2- He can't touch the ceiling because he is not tall \_\_\_\_\_ .  
 a. enough    b. yet    c. rather    d. ago
- 3- Would you mind \_\_\_\_\_ the blackboard ?  
 a. clean    b. to clean    c. cleaning    d. cleaned
- 4- Would you please \_\_\_\_\_ me wash the car ?  
 a. helping    b. help    c. to help    d. helped
- 5- I'd like \_\_\_\_\_ Suha .  
 a. to visit    b. visit    c. visiting    d. visited
- 6- They won't come \_\_\_\_\_ you invite them .  
 a. still    b. unless    c. therefore    d. for
- 7- He is very rich \_\_\_\_\_ he can buy a new car .  
 a. however    b. therefore    c. although    d. whatever
- 8- The book \_\_\_\_\_ his desk is mine .  
 a. on    b. with    c. above    d. out
- 9- If she helped me , I \_\_\_\_\_ help her .  
 a. shall    b. will    c. may    d. might
- 10- I want a pen . \_\_\_\_\_ ? To write a letter .  
 a. How long    b. When    c. How    d. What for
- 11- Shall we buy our sister a camera ?  
 \_\_\_\_\_ we don't have any money .  
 a. I agree    b. I'm afraid    c. I hope    d. I'd rather
- 12- \_\_\_\_\_ I borrow your umbrella ?  
 - Certainly .

- a. Should    b. Could    c. Would    d. Might
- 13- We go to school every day \_\_\_\_\_ holidays .  
a. during    b. since    c. until    d. except
- 14- She is the girl \_\_\_\_\_ mother is a nurse .  
a. whose    b. which    c. who    d. what
- 15- Nada said she \_\_\_\_\_ very tired .  
a. is    b. was    c. are    d. were
- 16- March is the \_\_\_\_\_ month of the year .  
a. first    b. next    c. third    d. last
- 17- Ali asked Huda if she \_\_\_\_\_ cook well .  
a. can    b. may    c. willd. could
- 18- Books are usually made of \_\_\_\_\_ .  
a. wool    b. wood    c. paper    d. metal
- 19- She has a \_\_\_\_\_ , She has to go to the dentist .  
a. headache    b. toothache    c. cold    d. flu
- 20- Gagarin was a \_\_\_\_\_ .  
a. teacher    b. baker    c. typist    d. pilot
- 21- I want to cash a cheque , I have to go to the \_\_\_\_\_ .  
a. theatre    b. bank    c. museum    d. factory
- 22- The sun gives us \_\_\_\_\_ .  
a. water    b. heat only    c. light only    d. heat and light
- 23- Iraq \_\_\_\_\_ oil .  
a. exports    b. imports    c. buys    d. makes
- 24- Thick clothes keep us \_\_\_\_\_ .  
a. sad    b. cold    c. warm    d. happy
- 25-The child is jumping \_\_\_\_\_ the window .  
a. out of    b. on    c. with    d. in front

26-My mother bought \_\_\_\_\_ car .

- a. a two-door    b. a two-doors    c. two-doors    d. two-door

27-The boy \_\_\_\_\_ in the garden is my brother .

- a. standing    b. stands    c. stood    d. is standing

28-What about \_\_\_\_\_ chess .

- a. playing    b. play    c. played    d. to play

29-The floor has \_\_\_\_\_ cleaned .

- a. been    b. is    c. was    d. be

***QII- Re-write the following sentences as in the example below :***

Example : They break the window .

The window is broken .

1. Huda helps her mother .
2. We use knives for cutting meat .
3. Iraq makes tractors .
4. We do not export ships .
5. Ali must post the letter soon .

***QIII- Write the correct form of the verb between brackets:***

1. He (clean) the blackboard yesterday before the teacher (come) in .
2. We (study) English since 1998 .
3. I (watch) television last night after I (write) the letter.
4. They (play) football for two hours .



***QIV- Answer the following questions in one continuous paragraph :***

Where did Nada go yesterday ? (to the station)

1- Why did she go there ? (to meet her father)

2- Who was with her ? (her mother)

3- Where had her father been ? (in Basrah)

4- How did they feel when they saw him ? (very happy)

***Q V- Read the following passage carefully :***

Yasir is a clever boy . He lives near the school , so he usually comes on foot . He always works hard before exams . His mother is proud of him because he has never failed . She wants him to be a doctor . " He can serve his country well " she says , and " as our country gives us every thing we want , we must give it whatever we could in return " . Good Iraqis should always be faithful to their country . They should do their best in every field they work in .

***Now write the number of the sentence and the letter of the correct answer :***

1. Yasir's mother wants him to be a \_\_\_\_\_ .

- a. pilot      b. teacher      c. doctor      d. butcher

2. Iraqi people should serve their \_\_\_\_\_ .

- a. friend      b. country      c. university      d. enemy

3. Yasir is a \_\_\_\_\_ .

- a. butcher      b. doctor      c. soldier      d. pupil

4. Yasir is \_\_\_\_\_ .

- A. clever      b. rich      c. kind      d. proud

5. He lives near the \_\_\_\_\_ .

- a. hospital   b. cinema   c. school   d. factory

**Q1**

Item No.	Suitable	Unsuitable	Suggested modification
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			

25			
26			
27			
28			
29			

**Q2**

Item No.	Suitable	Unsuitable	Suggested modification
1			
2			
3			
4			

## Appendix D

### The Final Version of the Achievement Test Given to the Pupils

***QI- Write the number of the sentence and the letter of most suitable choice :***

- 1- It is hot in Iraq \_\_\_\_\_ the summer .  
a. through    b. round    c. during    d. between
- 2- He can't touch the ceiling because he is not tall \_\_\_\_\_ .  
a. enough    b. yet    c. rather    d. ago
- 3- Would you mind \_\_\_\_\_ the blackboard ?  
a. clean    b. to clean    c. cleaning    d. cleaned
- 4- Would you please \_\_\_\_\_ me wash the car ?  
a. helping    b. help    c. to help    d. helped
- 5- I'd like \_\_\_\_\_ Suha .  
a. to visit    b. visit    c. visiting    d. visited
- 6- They won't come \_\_\_\_\_ you invite them .  
a. still    b. unless    c. therefore    d. for
- 7- He is very rich \_\_\_\_\_ ha can buy a new car .  
a. however    b. therefore    c. although    d. whatever
- 8- The book \_\_\_\_\_ his desk is mine .  
a. on    b. with    c. above    d. out
- 9- If she helped me , I \_\_\_\_\_ help her .  
a. shall    b. will    c. may    d. might
- 10- I want a pen .  
\_\_\_\_\_ ?  
To write a letter .  
a. How long    b. When    c. How    d. What for
- 11- Shall we buy our sister a camera ?  
\_\_\_\_\_ we don't have any money .  
a. I agree    b. I'm afraid    c. I hope    d. I'd rather

- 12- \_\_\_\_\_ I borrow your umbrella ?  
- Certainly .  
a. Should b. Could c. Would d. Might
- 13- We go to school every day \_\_\_\_\_ holidays .  
a. during b. since c. until d. except
- 14- She is the girl \_\_\_\_\_ mother is a nurse .  
a. whose b. which c. who d. what
- 15- Nada said she \_\_\_\_\_ very tired .  
a. is b. was c. are d. were
- 16- March is the \_\_\_\_\_ month of the year .  
a. first b. next c. third d. last
- 17- Ali asked Huda if she \_\_\_\_\_ cook well .  
a. can b. may c. will d. could
- 18- Books are usually made of \_\_\_\_\_ .  
a. wool b. wood c. paper d. metal
- 19- She has a \_\_\_\_\_ , she has to go to the dentist .  
a. headache b. toothache c. cold d. flu
- 20- Gagarin was a \_\_\_\_\_ .  
a. teacher b. baker c. typist d. pilot
- 21- I want to cash a cheque , I have to go to the \_\_\_\_\_ .  
a. theatre b. bank c. museum d. factory
- 22- The sun gives us \_\_\_\_\_ .  
a. water b. heat only c. light only d. heat and light
- 23- Iraq \_\_\_\_\_ oil .  
a. exports b. imports c. buys d. makes
- 24- Thick clothes keep us \_\_\_\_\_ .  
a. sad b. cold c. warm d. happy
- 25- The child is jumping \_\_\_\_\_ the window .  
a. out of b. on c. with d. of
- 26- My mother bought \_\_\_\_\_ car .  
a. a two-door b. a two-doors c. two-doors d. two-door

- 27- The boy \_\_\_\_\_ in the garden is my brother .  
 a. standing b. stands c. stood d. is standing
- 28- What about \_\_\_\_\_ chess .  
 a. playing b. play c. played d. to play
- 29- The floor has \_\_\_\_\_ cleaned .  
 a. been b. is c. was d. be

***QII- Re-write the following sentences as in the example below :***

Example : They break the window .  
 The window is broken .

1. Huda helps her mother .
2. We use knives for cutting meat .
3. Iraq makes tractors .
4. We do not export ships .
5. Ali must post the letter soon .

***QIII- Write the correct form of the verb between brackets:***

- 1- He (clean) the blackboard yesterday before the teacher (come) in .
- 2- We (study) English since 1998 .
- 3- I (watch) television last night after I (write) the letter.
- 4- They (play) football for two hours .

***QIV- Answer the following questions in one continuous paragraph :***

- 1- Where did Nada go yesterday ? (to the station)
- 2- Why did she go there ? (to meet her father)
- 3- Who was with her ? (her mother)
- 4- Where had her father been ? (in Basrah)
- 5- How did they feel when they saw him ? (very happy)

**Q V- Read the following passage carefully :**

Yasir is a clever boy . He lives near the school , so he usually comes on foot . He always works hard before exams . His mother is proud of him because he has never failed . She wants him to be a doctor . " He can serve his country well " she says , and " as our country gives us every thing we want , we must give it whatever we could in return " . Good Iraqis should always be faithful to their country . They should do their best in every field they work in .

**Now write the number of the sentence and the letter of the correct answer :**

- 1- Yasir's mother wants him to be a \_\_\_\_\_ .  
a. pilot    b. teacher    c. doctor    d. butcher
- 2- Iraqi people should serve their \_\_\_\_\_ .  
a. friend    b. country    c. university    d. enemy
- 3- Yasir is a \_\_\_\_\_ .  
a. butcher    b. doctor    c. soldier    d. pupil
- 4- Yasir is \_\_\_\_\_ .  
A. clever    b. rich    c. kind    d. proud
- 5- He lives near the \_\_\_\_\_ .  
a. hospital    b. cinema    c. school    d. factory

**Appendix E**  
**Subjects' Scores in Test-Retest for Estimating**  
**Reliability**

No.	X Test	Y Re-Test	XY	X <sup>2</sup>	Y <sup>2</sup>
1	18	17	306	324	289
2	20	22	440	400	484
3	19	18	342	361	324
4	17	20	340	289	400
5	10	12	120	100	144
6	12	12	144	144	144
7	16	20	320	256	400
8	20	17	340	400	289
9	22	15	330	484	225
10	21	24	504	441	576
11	15	18	270	225	324
12	23	20	460	529	400
13	26	22	572	676	486
14	24	26	624	576	676
15	12	14	168	144	196
16	31	27	837	961	729
17	28	24	672	784	576
18	15	18	270	225	324
19	17	11	187	289	121
20	19	22	418	361	486



21	20	29	580	400	841
22	29	30	870	841	900
23	16	19	304	256	361
24	17	20	340	289	400
25	18	22	396	324	484
26	9	14	126	81	196
27	10	18	180	100	324
28	12	16	192	144	256
29	25	28	700	625	784
30	32	27	864	1024	729
31	33	23	759	1089	529
32	16	18	288	256	324
33	17	22	374	289	484
34	20	22	440	400	484

**Appendix F**  
**The Scores of the Experimental and Control Groups**  
**in the Final Achievement Test**

The Experimental Group				The Control Group			
No.	Scores	No.	Scores	No.	Scores	No.	Scores
1	47	20	46	1	12	20	25
2	37	21	30	2	36	21	19
3	45	22	34	3	30	22	16
4	35	23	52	4	31	23	24
5	25	24	34	5	20	24	20
6	32	25	28	6	27	25	23
7	31	26	28	7	30	26	23
8	37	27	44	8	22	27	17
9	39	28	42	9	38	28	23
10	39	29	23	10	32	29	24
11	37	30	20	11	27	30	19
12	35	31	44	12	19	31	34
13	35	32	27	13	21	32	20
14	26	33	47	14	33	33	14
15	35	34	24	15	17	34	35
16	34	35	45	16	28	35	32
17	29			17	32		
18	49			18	41		
19	30			19	18		

## Appendix G

The Scores of the Final Achievement Test in : Structure, Vocabulary, Composition and Reading Comprehension

No.	Experimental Group				Control Group			
	STR	VOC	COM	R. C.	STR	VOC	COM	R. C.
1	30	6	7	4	5	2	3	2
2	21	6	5	5	23	5	6	2
3	32	5	5	3	18	4	5	3
4	17	6	7	5	18	3	6	4
5	14	2	4	5	10	2	5	3
6	16	3	8	5	14	3	7	3
7	15	3	8	5	17	4	6	3
8	19	6	7	5	11	2	6	3
9	21	5	9	4	20	6	7	5
10	18	7	9	5	17	4	7	4
11	22	5	6	4	15	3	6	3
12	17	6	7	5	10	3	4	2
13	17	4	9	5	12	5	0	4
14	11	3	8	4	21	4	6	2
15	18	6	7	4	11	2	2	2
16	17	5	9	3	15	3	5	5
17	17	4	3	5	19	3	6	4
18	31	6	7	5	25	5	7	4
19	16	4	6	4	6	3	4	5
20	30	5	8	3	9	3	8	5
21	14	3	8	5	8	5	6	0
22	17	4	8	5	5	4	3	4
23	32	7	8	5	13	3	5	3
24	18	5	7	4	12	3	2	3
25	17	3	5	3	11	4	6	2
26	13	4	7	4	9	5	5	4
27	27	6	7	4	8	3	4	2
28	24	5	8	5	10	5	5	3
29	13	3	4	3	10	5	5	4
30	10	4	4	2	9	3	5	2
31	29	5	6	4	25	3	3	3
32	16	3	5	3	7	5	6	2

33	29	6	7	5	6	0	5	3
34	11	4	6	3	22	4	6	3
35	30	6	4	5	19	4	5	4

## List of Abbreviations

Str.	Structure
Voc.	Vocabulary
Com.	Composition
RC	Reading Comprehension
NECI	New English Course for Iraq

# List of Appendices

No.	Content	Page
A	A Model Lesson Plan for the Control Group	76
B	The Objectives of the Programme in the Intermediate Stage	80
C	The Initial Version of the Achievement Test given to the Jury	81
D	The Final Version of the Achievement Test given to the pupils	89
E	Subjects' Scores in Test-Retest for Estimating Reliability	93
F	The Scores of the Experimental and Control Groups in the Final Achievement Test	95
G	The Scores of the Two Groups in the Final Achievement Test in : Structure , Vocabulary , Composition and Reading Comprehension	96

# List of Figures

*List of Figures*

No.	Title	Page
1	The continuous Teaching – Learning Process	3
2	A Criterion – referenced teaching model	17
3	A Norm – referenced teaching model	17

# List of Tables

No.	Content	Page
1	The distribution of the subjects of the study	40
2	Chi-square and other statistics for the level of fathers' education	42
3	Chi-square and other statistics for the level of mothers' education	43
4	t-value and other statistics for the Age of the subjects	43
5	t-value and other statistics for the pupils' level of Achievement in English in the first intermediate Grade	44
6	t-value and other statistics for the pupils' level of Achievement in English in the first month of the second intermediate Grade	45
7	Time Allocation to the Various Activities	50
8	The difficulty level and the discrimination power of the items	63
9	t-test statistics for the study groups' achievement in structure	67
10	t-test statistics for the study groups' achievement in vocabulary	68
11	t-test statistics for the study groups' achievement in composition	69
12	t-test statistics for the study groups' achievement in reading comprehension	69
13	t-test statistics for the study groups' achievement in entire achievement	70



## ***Bibliography***

- Ahmann , J. Stanley , and Marvin D. Glock . (1975) . *Measurement and Evaluating Educational Achievement* . 2<sup>nd</sup> ed. Boston : Allyn and Bacon , Inc.
- Alderson , J. Charles , et al., (1995) . *Language Test Construction and Evaluation* . Cambridge : Cambridge University Press .
- Al-Hamash , Khalil I. (1984) . *A Survey of English Textbooks in Primary and Secondary Schools in Iraq* . Baghdad : Al-Rashid Press .
- Al-Hamsh , Khalil I., A. J. Al-Jubouri , & W. M. Al-Hiti . (1989) . *Testing Guide for Primary and Intermediate Teachers of English in Iraq* . 5<sup>th</sup> revised ed. Arbeel : Ministry of Education Press .
- Al-Hamash , Khalil I., Ra'ad Ahmed , & Khudheyer S. Al-Khazraji . (1990) . *The New English Course for Iraq , Book 4 : Teacher's Guide* . 2<sup>nd</sup> revised ed. Baghdad : Al- Mansour Press Company .
- Al-Sa'doon , Adila Ali .(2000). *The Impact of Formative Evaluation Strategy in Achievement of Second Intermediate Girl Student in Islamic Education* . Unpublished Master Thesis . University of Baghdad .
- Applebee , Arthur . (1988) . “ *The Enterprise We are part of : Learning to Teach* ” . In *The Word for Teaching is Learning* , edited by M. L. Lightfoot and N. Martin . London : Heinemann Educational Books .
- Bachman , Lyle F. and Adrian S. Palmer . (1996) . *Language Testing in Practice* . Oxford : Oxford University Press .
- Block , J. H. (1971) . *Mastery Language Theory and Practice* . 2<sup>nd</sup> ed. New York : Holt , Rinehart and Winstone .
- Bloom , B. S. Hasting J. T., & Madaus G. F. (1971) . *Handbook on Formative and Summative Evaluation of Student Learning* . New York . McGraw-Hill .
- Bloom , B. S. (1976) . *Human Characteristics and School Learning* . New York : McGraw-Hill Book Company .

- Bloom , B. S., G. F. Madaus , & J. T. Hastings . (1981) . *Evaluation to Improve Learning* . New York: McGraw-Hill Book Company .
- Bronghton , G., C. Brumfit , R. Flavell , P. Hill , and A. Princae . (1981). *Teaching English as a Foreign Language* . 2<sup>nd</sup> ed. London : Routledge & Kegan Paul .
- Brown , F. G. (1976) . *Principles of Educational and Psychological Testing* . 2<sup>nd</sup> ed. New York : Holt , Rinehart and Winston .
- Callahan , Joseph F. and Clark , Leonard H. (1983) . *Foundations of Education Planing of Competence* . 2<sup>nd</sup> ed. London : Collier Macmillan Publishers . London .
- Campbell , Donald T., and Julian C. Stanley . (1963) . *Experimental and Quasi-Experimental Designs for Research* . Boston : Houghton Mifflin Company .
- Carey , L., & Dick W., (1981) . *The Systematic Design of Instruction* . Dallas Texas : Scott Foresman Co.
- Carol , Puhl , A. (1997) . *Continuos Assessment in ESL Classroom* . In Forum Journal Vol. 35 No. 2 April 1997 .
- Christensen , Larry B. (1980) . *Experimental Methodology* . 2<sup>nd</sup> ed. Boston : Allyn and Bacon , Inc.
- Darwash , A. Ali & Faris Al-Jarah (1989) . *An Eliminary Course in Testing English as a Foreign Language for Student Teachers* . 3<sup>rd</sup> ed. Baghdad : Al-Adeeb Press .
- Downi , N. M. (1967) . *Fundamental of Measurement : Technique and Practice* . 2<sup>nd</sup> ed. London : Oxford University Press .
- Dwyer , Carol .A. (1982) . “ *Achievement Testing* ” . In *Encyclopedia of Educational Research , Vol.1* , edited by H. E. Mitzel . 5<sup>th</sup> ed. New York : The Free Press , Macmillan Publishing Co., Inc.
- Ebel , Robert L., (1979) . *Essentials of Educational Measurement* . 3<sup>rd</sup> ed. New Jersey : Englewood Cliffs . Prentice-Hall .
- Gipps , Caroline V. (1994) . *Beyond Testing : Towards a Theory of Educational Assessment* . 1<sup>st</sup> ed. London : The Falmer Press .

- Glaser , R. (1963) . Instructional Technology and The Measurement of Learning outcomes : Some Questions . *International Journal of Educational Research* , 14-5 .
- Glass , Gene V. & Stanley , Julian C. (1970) . *Statistical Methods in Education and Psychology* . New Jersey : Prentice-Hall Inc.
- Good , Carter V., ed. (1973) . *Dictionary of Education* . 3<sup>rd</sup> ed. New York : McGraw-Hill Book Co.
- Goodson , M. L. (1975) The Effect of Objective-based diagnostic Test and help sessions on the Achievement of undergraduate physical science students . (Doctoral Dissertation , Indiana University) . *Dissertation Abstract International* , Vol. 36 , No.9 , March. 1976. pp. 5969-5970 .
- Gronlund , N. E. (1965) . *Measurement and Evaluation in Testing* . New York : Macmillan Publishing Co., Inc.
- \_\_\_\_\_ (1976). *Measurement and Evaluation in Teaching* . 3<sup>rd</sup> ed. New York : Macmillan Publishing Co., Inc.
- \_\_\_\_\_ (1977) . *Constructing Achievement Tests* . 2<sup>nd</sup> ed. Englewood Cliffs , N. J. : Prentice-Hall , Inc.
- \_\_\_\_\_ (1982) . *Constructing Achievement Tests* . 3<sup>rd</sup> ed. Englewood Cliffs : Prentice-Hall , Inc.
- Harris , David P. (1969) . *Testing English as a Second Language* . New York : McGraw-Hill Book Company .
- Harrison , Andrew . (1983) . *A Language testing Handbook* . London: The Macmillan Press Limited .
- Heaton , J. B. (1975) . *Writing English Language Tests* . 1<sup>st</sup> ed . Essex : Longman .
- Hughes , A. (1989) . *Testing for Language Teachers* . Cambridge : Cambridge University Press .
- Karmel L. J., (1977) . *Measurement and Evaluation in the Schools* . New York : Macmillan Publishing Co., Inc.

- Kerlinger , F. N. (1978) . *Foundations of Behavioural Research* . 2<sup>nd</sup> ed.  
New York : Macmillan Publishing Co., Inc.
- Knight .D. J., (1983) . The Effect of Diagnostic Testing on The Achievement in Mathematics of Junior Grade Students (Doctoral Dissertation , University of Toronto . Canada) . In *Dissertation Abstract International* , Vol. 45, No. 2, August 1984, pp. 499-500-A.
- Lado , Robert . (1972) . *Language Testing : The Construction and Use of Foreign Language Tests* . 7<sup>th</sup> Impression . London : Longman .
- Lewin , Miriam . (1979) . *Understanding Psychological Research* .  
New York : John Wiley & Sons .
- Madsen , Harold S. (1983) . *Techniques in Testing* . London : Oxford University Press .
- Margot R., (1982) . *The Effect of Formative Assessment and Correctives on Learning Achievement* .(Doctoral Dissertation , University of South California) . In *Dissertation Abstract International* . Vol. 45, No. 3, June 1983, p. 127-A.
- Mouly , G. G. (1978) . *Educational Research : The Art and Science Investil – gatio* . Boston : Allyn and Bacon . Inc.
- Muheidat , Abed Al-Hakeem . (1990) . *The Effect of Formative Testing Strategy Upon Students' Achievement in Math* . Unpublished Master Thesis . Jordan University , Amman .
- Nader , Thabet Mohammad . (1999) . *The Attitudes of the English Language Toward Using Formative Evaluation in Improving EFL Learning in Tulkarm District* . Jordon : Athesis Submitted to al-Najah National University .
- Oller , John W. (1979) . *Language Tests at School . A Pragmatic Approach* . London Group Ltd. : Longman .
- Page , G. Terry , and J. B. Thomas and Marshal A. R. (1977) . *International Dictionary of Education* . Boston : The MIT Press .

- Patton M. Q., (1987) . *How to use qualitative Methods in English* . San Francisco : Sage Publications Inc.
- Payne , David A. (1974) . *The Assessment of Learning Cognitive and Affective* . University of Georgia , D. C. Heath and Company .
- Popham , W. James . (1975) . *Educational Evaluation* . Englewood Cliffs N.J : Prentice-Hall , Inc.
- Popham , W. James . (1993) . *Educational Evaluation* . 3<sup>rd</sup> ed. London: Allyn and Bacon .
- Remmers , H. H., N. L. Gage , and J. F. Rummel . (1965) . *A Practical Introduction to Measurement and Evaluation* . 2<sup>nd</sup> ed. New York : Harper & Ram , Publishers .
- Rivers , Wilga M. (1972) . *The Psychologist and the Foreign – Language Teacher* . 8<sup>th</sup> impression . Chicago : The University of Chicago Press .
- Robinson , Paul W. (1981) . *Fundamental of Experiment Psycology* . 2<sup>nd</sup> ed. Englewood Cliffs , N. j. : Prentice-Hall , Inc.
- \_\_\_\_\_ . (1976) . *Fundamental of Experimental Psychology: A Comparative Approach* . Englewood Cliffs , N., J. : Printice-Hall . Inc.
- Roscoe , John T. (1969) . *Fundamental Research Statistics (for the Behavioural Sciences)* . New York : Holt , Rineheart and Winston , Inc.
- Sadler , R. (1989) . *Formative Assessment and the Design of Instructional System* . New York : John Wiley & Sons .
- Scannell , Dale P., and D. B. Tracy . (1975) . *Testing and Measurement in the Classroom* . Boston : Houghton Mifflin Company .
- Thomas D. A., (1983) . The Effect of Formative Testing . Prescribed remediation , and retesting on student Performance in Calculus (Doctoral Dissertation , Montana State University) . In Dissertation Abstract International , Vol. 44, No. 8, February 1984, p. 2398-A .

- Thorndike , Robert & Elizabeth Hagen . (1969) . Measurement and Evaluation in Psychology and Education . 3<sup>rd</sup> ed. New York : John Wiley and Sons , Inc.
- Tuckman , Bruce W. (1975) . Measurement Educational Outcomes : Fundamentals of Testing . New York : Harcourt Brace Jovanovich , Inc.
- Tyler , Ralph W. (1969) . Basic Principles of Curriculum and Instruction. 29<sup>th</sup> impression . Chicago : The University of Chicago Press .
- Vallette , Rebecca M. (1967) . Modern Language Testing : A handbook . New York : Harcourt , Brace & World , Inc .

**المصادر العربية :**

١- الزوبعي ، عبد الجليل و الغنام ، محمد أحمد . (١٩٨١) . **مناهج البحث في التربية و علم النفس ، الجزء الأول . مطبعة جامعة بغداد .**

٢- فان دالين ، ديوبولدب . (١٩٨٠) . **مناهج البحث في التربية و علم النفس . ترجمة محمد نبيل نوفل و آخرون ط٣ ، القاهرة : مكتبة الأنجلو المصرية .**

## Table of Contents

	<b>Pages</b>
Dedication	V
Acknowledgements	VI
Abstract	VII
List of Abbreviations	IX
List of Tables	XIII
List of Figures	XIV
List of Appendices	XV
Chapter One : Introduction	
1.1 The Problem	1
1.2 Aims of Study	7
1.3 Hypothesis	8
1.4 Significance of the Study	8
1.5 Plan of the Study	9
1.6 Limits of the Study	9
1.7 Definition of Basic Terms	10
Chapter Two : Theoretical Background	
2.1 An Introductory Note	12
2.2 Formative Assessment Versus Summative Assessment	13
2.3 Testing	18
2.4 Types of Tests	19
2.4.1 Types of Test according to Function	20
2.4.2 Types of Test according to the Stages of instruction	21
2.4.3 Types of Test according to Contents	22
2.5 Formative Test Construction	23
2.6 Test Qualities	23
Chapter Three : Review of Previous Studies	
3.1 An Introductory Note	27
3.1.1 Goodson (1975)	27
3.1.2 Margot (1982)	29
3.1.3 Knight (1983)	29



3.1.4 Thomas (1983)	30
3.1.5 Muheidat (1990)	31
3.1.6 Nader (1999)	32
3.1.7 Al-Sa'doon (2000)	33
3.2 Discussion of the Previous Studies	33
Chapter Four : Procedures	
4.1 An Introductory Note	37
4.2 The Experimental Design	37
4.3 Sample Selection Procedures	39
4.3.1 Fathers' Level of Education	41
4.3.2 Mothers' Level of Education	42
4.3.3 The Age of Subjects	43
4.3.4 The Subjects' Level of Achievement in English in the First Intermediate Grade	44
4.3.5 Pupils' Scores in English in the First Month	44
4.3.6 The Extraneous Factors Jeopardizing External	45
4.3.6.1 History	45
4.3.6.2 Maturation	46
4.3.6.3 Selection Bias	46
4.3.6.4 Experimental Mortality	46
4.4 Instrumentation	47
4.5 The Instructional Materials	47
4.6 The Experiment Application	49
4.6.1 The Control Group	50
4.6.2 The Experimental Group	51
4.6.3 Instruments of the Research	52
4.6.4 The Achievement Test	52
4.6.5 Test Construction	53
4.7 The Written Test Validity	54
4.8 The Pilot Study	57
4.9 The Test Reliability	57
4.10 Test Administration	59
4.10.1 The Scoring Scheme	60
4.10.2 Item Analysis	61

4.10.3 Statistical Tools	65
Chapter Five : Results	
5.1 An Introductory Note	66
5.2 Achievement in Specific Components of the Test	66
5.2.1 Achievement in Structure	66
5.2.2 Achievement in Vocabulary	67
5.2.3 Achievement in Composition	68
5.2.4 Achievement in Reading Comprehension	69
5.3 The Entire Achievement Test	70
5.4 Interpretation	71
Chapter Six : Conclusions , Recommendations and Suggestions for Further Research	
6.1 Conclusions	73
6.2 Recommendations	74
6.3 Suggestions for Further Research	75
Appendices	76
Bibliography	98
Arabic References	104
Abstract in Arabic	٢-١

We certify that this dissertation has been prepared under our supervision at the University of Baghdad as a partial fulfilment of the requirements for the degree of Doctor of Philosophy in Education (Teaching English as a Foreign Language) .

Signature :

Name : Prof. Dr.

Abdullah H. Al-Musawi

Signature :

Name : Asst. Prof. Dr.

Abdul-Latif Al-Jumaily

In view of the available recommendation , I forward this dissertation for debate by the examining committee .

Signature :

Name :

Chairman of the Department of Educational and Psychological Sciences .

Date :     /     /2004

We certify that we have read the dissertation entitled “ The Effect of Formative Assessment on the Pupils’ Achievement in English ” by *Ridha Ghanim Dakhil Al-Saedi* , and as an examining committee examined the student in its content and that, in our opinion , it is adequate as a dissertation for the degree of Doctor of Philosophy in Methods of Teaching English as a Foreign Language .

Signature :

Name: Prof. Dr.

Muna Younis Bahri

Member

Signature :

Name: Asst. Prof. Dr.

Riyadh T. Kadhim

Member

Signature :

Name: Asst. Prof. Dr.

Ilham N. Khurshid

Member

Signature :

Name: Asst. Prof. Dr.

Mu'ayad M. Sa'eed

Member

Signature :

Name : Prof. Ayif Habib Al-Ani

Chairman

Approved by the Council of the College of Education/ Ibn Rushd,  
University of Baghdad .

Signature :

Name : Dr. Abdul-Amir A. Dixon .

Dean of the College of Education / Ibn Rushed , University of  
Baghdad .

Date :        /        /2004

بسم الله الرحمن الرحيم

جمهورية العراق



وزارة التعليم العالي و البحث العلمي  
جامعة بغداد  
كلية التربية / ابن رشد  
الدراسات العليا

العدد :

التاريخ :

أمر إداري

**تشكيل لجنة مناقشة أطروحة طالب الدكتوراه**

تقرر تشكيل لجنة من السادة المدرجة أسماؤهم في أدناه لمناقشة طالب الدكتوراه (رضا غانم داخل الساعدي) عن أطروحته الموسومة بـ (أثر التقويم التكويني في تحصيل الطلبة في اللغة الإنكليزية) في قسم العلوم التربوية و النفسية في الساعة العاشرة من صباح يوم السبت ٢٦/٦/٢٠٠٤ في قاعة المرحوم الدكتور ناظم العبيدي .

ت	الاسم و اللقب العلمي	التخصص	مكان العمل	الصفة
١	أ. عايف حبيب العاني	مناهج و طرائق تدريس	جامعة بغداد/كلية التربية/ابن رشد	رئيساً
٢	أ. د. منى يونس بحري	مناهج و طرائق تدريس	جامعة بغداد/كلية التربية/ابن رشد	عضواً
٣	أ. م. د. رياض طارق العميدي	طرائق تدريس اللغة الإنكليزية	جامعة بابل/كلية التربية	عضواً
٤	أ. م. د. الهام نامق خورشيد	طرائق تدريس اللغة الإنكليزية	جامعة ديالى/كلية التربية	عضواً
٥	أ. م. د. مؤيد محمد سعيد	طرائق تدريس اللغة الإنكليزية	جامعة بغداد/كلية التربية/ابن رشد	عضواً
٦	أ. د. عبد الله حسن الموسوي	مناهج و طرائق تدريس	جامعة بغداد/كلية التربية/ابن رشد	مشرفاً
٧	أ. م. د. عبد اللطيف علوان الجميلي	اللغة الإنكليزية	جامعة بغداد/كلية الآداب	مشرفاً

أ. د. طارق نافع الحمداني

معاون العميد للشؤون العلمية و الدراسات العليا

نسخة منه إلى/

رئاسة جامعة بغداد / قسم الدراسات العليا ... للتفضل باتخاذ ما يلزم ، و نرفق طياً نسخة من أطروحة الطالب أعلاه .  
 القسم المالي / ... لصرف مبلغ (٢٥٠٠٠) خمسة و عشرون ألف دينار لكل عضو مناقش و (٥٠٠٠٠) خمسون ألف دينار للمشرف على الأطروحة ، و صرف مبلغ مكافأة (٢٥٠٠٠) خمسة و عشرون ألف دينار للخبير اللغوي د. مؤيد محمد سعيد ، و صرف مبلغ مكافأة (٢٥٠٠٠) خمسة و عشرون ألف دينار للخبير العلمي أ. م. د. ناهدة طه الناصري .

الدراسات العليا / ملفه الطالب .

قسم العلوم التربوية و النفسية .

السادة رئيس و أعضاء لجنة المناقشة مع نسخة من أطروحة الطالب أعلاه ... مع التقدير .