An-Najah National University

College of Education

The Effect of Question's Types and Levels

on Students' Academic Achievement

Master Thesis

Rafi' A. Daraghmeh

Supervised by

Prof. Afnan Darwazeh

In Partial Fulfillment of Requirements for the Degree

of Master of Education

Approved by :

Prof. Afnan Darwazeh   /An-Najah / (Chairman)

Dr. Suzanne Arafat   /An-Najah /   (Member)

Dr. Nabil Alawi   /An-Najah /   (Member)

Dr. Hanna Tushyeh   /Bethlehem /   (External Member)

ACKNOWLEDMENT

I would like to express my deepest gratitude to my supervisor "Prof. Afnan Darwazeh" for her enthusiastic support and guidance to me from the first step of this study. Without her help this study would not have been the best of what I can produce.

My greatest thanks go as well to the members of the thesis "Dr. Suzanne Arafat", "Dr. Nabil Alawi", and "Dr. Hanna Tushyeh" who helped me with their guidance and comments. I pay tribute to Dr. Nabil Alawi, who helped and advised me all the time.

I dedicate my thesis to my parents, to my wife, my daughter - Du'a'-one year old, my son Ahmad - whose timely arrival on the day of defending this thesis made all the difference- and to all those who are interested in testing and evaluation.

I extend heartfelt thanks to all my teachers at An-Najah National University, especially Dr. Odeh Odeh, Dr. Roqaya Hirzallah, Dr. Rami Hamdallah, Dr. Fawaz Aqel and all who helped me in this research..

I owe special thanks to Dr. Jein Qattan from Bethlahem University for her encouragement and help.

Finally, I owe a debt of gratitude to my respective parents for their continuous help and encouragement.

Rafi' A. Daraghmeh

Nablus, October, 1997

ABSTRACT

The present study aimed at investigating the effect of question's types (essay vs. multiple-choice), question's levels (remember - an - instance RI, remember- a - generality RG, and Use - a - generality UG), in one hand and Student's ability (high, medium, and low), on the other hand, on students' academic achievement.

To accomplish this aim, a random sample consisted of one-hundred and forty freshman university students at An-Najah National University were presented with a 1159 - word passage chosen from their required university textbook University English I. The sample of this study was randomly assigned into seven groups: Six groups were given the passage followed with post-passage questions. Three groups were exposed to multiple-choice questions written on three levels of learning according to Merrill's Taxonomy - Remember - an - Instance (RI), Remember - a - Generality (RG), and Use - a - generality (UG). Three other groups were exposed to essay questions written on the same mentioned learning levels (RI, RG, and UG). One group was used as a control group which was given the passage only without any post-passage questions.

A 19-item posttest comprised of two types of questions (essay and multiple-choice) were written on the passage measuring all specific and general information on three levels of learning (RI, RG, and UG). The RI and RG posttest questions were the same as those which appeared in the passage during the experiment. The UG posttest questions were not the same as those which appeared in the passage during the experiment, otherwise they will measure the remember of application level rather than the application level itself.

Three-Way Analysis of Variance (ANOVA) design 2x3x3 (types of questions: essay and multiple-choice, levels of questions: RI, RG, and UG, and students ability: high, medium, and low) was used for analyzing the data of the study by using 'F' test. One Way Analysis of Variance (1x7) was also used to compare the six experimental groups with the control group. "Scheffe" test was used whenever the general 'F' test revealed significant differences on .05 level.

The major results that the researcher found significant on some sub-tests were the following :
- On the remember - an - instance   multiple-choice sub-test, there was a significant difference main effect for question levels which

indicated that RG question level was better than UG and RI question levels respectively.

- On the remember - an - instance essay sub-test, there was a significant main effect for question types which indicated that multiple-choice type had a greater effect on student's academic achievement than essay question type.

- On the remember - a - generality multiple-choice sub-test, there was a significant main effect for question types which indicated that essay type had a greater effect on student's academic achievement than the multiple-choice type.

- On the use - a - generality essay sub-test, there was a significant main effect for question levels which indicated that RG question level had a greater effect on student's academic achievement than UG and RI question levels respectively.

But on the overall learning, the directions of the means showed the following results, though it was no significant difference:

1. Students who received remember - a - generality (RG) then who received use - a - generality (UG) level questions performed better   than their colleagues who received remember - an - instance (RI)level questions.
2. Students who received essay questions performed better than students who received the multiple-choice questions.
3. The performance of the experimental groups who were exposed to post-passage questions was higher than the   performance of the control group who were not exposed to any type of question during experiment.
4. High student's ability performed significantly better than the medium then the low student's ability.

Based on these results, the researcher   recommended teachers:1) to use the RG level questions frequently in order to help students to understand the intended text and then to answer the higher level of questions like the application level. 2) to use the essay type questions frequently in order to promote students' cognitive strategies deeply, hence to increase their academic achievement on all levels of learning.

الملخص

أثر نمط الأسئلة التعليمية ومستواها على تحصيل الطالب الأكاديمي

هدفت الدراسة إلى بحث أثر نوع السؤال التعليمي(مقالي / اختيار من متعدد)، ومستواه (تذكر المثال او الحقيقة، تذكر معلومات عامة، تطبيق معلومات عامة)، وقدرة الطالب وفق معدله في امتحان الثانوية العامة (عليا، متوسطة، دنيا) على تحصيل الطالب الأكاديمي.

و لتحقيق هذا الهدف قام الباحث باختيار عينة عشوائية تكونت من (١٤٠) طالبا من طلبة السنة الاولى في جامعة النجاح وتعريضهم لقراءة نص تعليمي تضمن (١١٥٩) كلمة اخذت من كتاب "اللغة الانجليزية ١ " المقرر في الجامعة. وزعت العينة عشوائياً الى سبع مجموعات: ستة مجموعات تجريبية تلقت النص التعليمي متبوعاً بمجموعة من الأسئلة وضعها الباحث وفق نظرية "دافيد ميرل" (١٩٨٣) للاهداف التربوية ، وهي كما يلي: ثلاث مجموعات تجريبية تلقت اسئلة من نوع اختيار من متعدد كتبت على ثلاثة مستويات من التعلم: مجموعة تلقت اسئلة تقيس مستوى تذكر المثال او الحقيقة، ومجموعة تلقت اسئلة تقيس مستوى تذكر المعلومات العامة، ومجموعة ثالثة تلقت اسئلة تقيس مستوى تطبيق المعلومات العامة) .اما المجموعات الثلاث الاخرى فقد تلقت الاسئلة نفسها ولكن بنمط مقالي، في حين تلقت المجموعة الضابطة النص التعليمي فقط دون اي نوع من الاسئلة.

وضع الباحث اختبار تحصيلي لاحق تألف من (١٩) فقرةً قاست مستوى تذكر المثال والحقيقة، ومستوى تذكر المعلومات العامة ومستوى تطبيق المعلومات العامة.

استخدم الباحث تحليل التباين الثلاثي (٣x ٣x٢) "Three-Way Analysis of Variance" ANOVA لتحليل بيانات الدراسة بواسطة اختبار "ف" F-test ، حيث يمثل العامل الاول نوع السؤال التعليمي (مقالي / اختيار من متعدد)،ويمثل العامل الثاني مستوى السؤال التعليمي (تذكر المثال او الحقيقة، تذكر معلومات عامة، تطبيق معلومات عامة)، في حين يمثل العامل الثالث قدرة الطالب وفق معدله في امتحان الثانوية العامة (عليا، متوسطة، دنيا). كما أستخدم تحليل التباين الاحادي (٧x ١) "One Way Analysis of Variance" لمقارنة المجموعات التجريبية مع المجموعة الضابطة. واستخدم اختبار شيفي "Scheffe" في تحليل التباين اللاحق للكشف عن مكان الفروق بين المتوسطات فيما اذا كان اختبار "ف" له دلالة احصائية.

اظهرت النتائج بعض الفروقات الاحصائية على بعض الاختبارات الفرعية (Sub-tests) حيث وجد

* فرق احصائي على الاختبار الفرعي الموضوعي الذي قاس **تذكر المثال او الحقيقة** (RIM) لصالح مستوى الاسئلة التعليمية مفاده ان مستوى التذكر العام RG كان قد حسن

اداء الممتحنين بشكل افضل من مستوى تذكر المثال او الحقيقة RI، ومستوى التطبيق العـام UG.

*   ووجـد ايضـاً فرق احصـائي على الاختبـار الفرعي الموضـوعي الـذي قـاس **تـذكر المعلومات العامة** RGM لصالح نمط الاسئلة التعليمية مفاده ان النمط المقـالي كان لـه اثر اكبر في تحسين اداء الممتحنين من نمط الاختيار من متعدد.

*   هناك فرق احصائي على الاختبار الفرعي المقالي الـذي قـاس **تذكر المثـال او الحقيقة** (RIE) لصالح نمط الاسئلة التعليمية مفاده ان نمط الاختيـار من متعدد كـان لـه اثر اكبر في تحسين اداء الممتحنين من النمط المقالي.

*   هنـاك فـرق احصـائي على الاختبـار الفرعي المقالي الـذي قـاس **تطبيق المعلومات العامة** (UGE) لصالح مستوى الاسئلة التعليمية مفاده ان مستوى التطبيق العام RG كان لـه اثر اكبر في تحسين اداء الممتحنين من مستوى تذكر المثـال او الحقيقة RI ، ومستوى تطبيق المعلومات العامة UG.

وبشكل عام فقد اظهرت الدراسة على الاختبار الكلي التوجه التالي في النتائج:

١. كان متوسط تحصيل الطلبة الذين تلقوا اسئلة على مستوى التذكر العـام اولا ثم مستوى تطبيق المعلومات العامة ثانيا اعلى من متوسط تحصيل الطلبة الذين تلقوا اسئلة على مستوى تذكر المثال او الحقيقة.

٢. كان متوسط تحصيل الطلبة الذين تلقوا اسئلة من النوع المقالي اعلى من متوسط الطلبة الذين تلقوا اسئلة من النوع الاختيار من متعدد.

٣. كان متوسط تحصيل الطلبة من ذوي القدرات العليا اولا ثم من ذوي القدرات المتوسطة ثانياً اعلى وبفرق احصائي من متوسط تحصيل زملائهم الطلبة من ذوي القدرات الدنيا.

٤. كان متوسط تحصيل المجموعات التجريبية الذين عالجوا اسئلة تعليمية بعد قراءتهم للنص التعليمي المدروس وفي اثناء التجربة اعلى من متوسط المجموعة الضابطة التي تلقت النص التعليمي فقط دون معالجة اية اسئلة.

وعليه يوصي الباحث بشكل رئيس المعلمين باستخدام الاسئلة المقالية اكثر من استخدام الاسئلة الموضوعية، واستخدام الاسئلة التي تقيس مستوى تذكر المعلومـات العامة اكثر مـن تلك التي تقيس مستوى تذكر المثـال او الحقيقة، اذ أن الاسئلة التي تقيس مستوى تذكر المعلومات العامة تساعدهم على فهم النص التعليمي ومن ثم الاجابة عن اي سؤال لاحق مهما كان المستوى العقلي الذي يقيسه.

Table of Contents

Chapters	Pages

List of Tables

List of Figures

Table of Appendixes

An-Najah National University

Collage of Education

The Effect of Question's Types and Levels

on Students' Academic Achievement

Master Thesis

Rafi' A. Daraghmeh

Supervised by

Prof. Afnan Darwazeh

In Partial Fulfillment of Requirements for the Degree

of Master of Education

Nablus, October, 1997

Chapter One

Chapter One

= = = = = = =

Introduction

= = = = = = =

Instructional assessment plays an essential role in the progress of the instructional process. It is an integral part of the teaching - learning process.

Students' learning can be measured by different procedures. One of these procedures, which is widely used by teachers in the classroom, is the achievement test. Achievement tests are commonly used for assessing students' learning. They are also considered as a tool for enabling teachers to judge whether students have mastered the taught skills and knowledge, thus to find out whether the planned behavioral objectives are achieved.

Educators (e.g., Brown, 1987; Darwazeh, 1997a) have classified achievement tests into two major types: the essay type, and the objective one. An essay type test is defined as a tool which requires students to write the answer in their own words (e.g., How do colds spread ?), whereas the objective type test is defined as a tool which requires students to recognize and mark the correct answer from a number of options (e.g., which disease of the following has the same symptoms to the cold ? a) Influenza b) Fever c) Cancer d) Heart disease).

Good achievement tests are supposed to measure different levels of learning. Bloom (1956), for example, classified six hierarchical categories according to the cognitive process they need. Descriptions of the major categories in the cognitive domain are: (1) knowledge levels in which students are required to memorize facts and knowledge (e. g., who was the winner of the 1954 Nobel prize for chemistry ?), (2) comprehension levels in which students are required to interpret information in their own words (e.g., why can't a doctor tell if a person has a cold or a flu ?), (3) application levels in which students are required to use what they have learned in a new situation (e.g., do you think that you should or should not shake hands with some one who has cholera ?), (4) analysis levels in which students are required to break down the instructional task into its components (e.g., what elements does water consist of ?), (5) synthesis levels in which students are required to make some inferences and come up with conclusions (e.g., what are the main ideas that the lesson talks about?), and (6) evaluation levels in which students are required to judge and value the merit of the learned materials, and make a decision (e.g., do you think that your teacher has explained the lesson of "The Common Cold" in terms of its major ideas clearly ?) (See, Darwazeh,1987, 1997a; Martin et al. 1994, p. 160-171; Hill,1982, pp. 181-200).

David Merrill (1983), on the other hand, has proposed different kinds of taxonomies based on two dimensions: 1) the type of instructional content, and 2) the level of instructional performance. The instructional content type is defined as all knowledge and information that are required by students to learn during their learning. Merrill (Ibid) had classified content type into four types.They are: (1) concepts, (2) principles, (3) procedures, and (4) facts (See, Darwazeh, 1995, for definition of each).

Whereas, the level of instructional performance is defined as the student's behavior which he /she shows after the instructional process has taken place. Merrill had classified the performance levels into four levels of learning. They are :

1- Remember - an - Instance level (RI) in which students are required to recall or recognize a specific information, such as to remember names, dates, symbols, labels, etc. (e.g., who was the winner of the 1954 Nobel prize for chemistry ?).

2- Remember - a - Generality level (RG) in which students are required to remember general information and basic ideas, such as recalling a definition of principle, concept, or procedure (e.g., what is the main difference between cold and influenza ?).

3- Use - a - Generality level (UG) in which students are required to apply a generality to a new instance (e.g., is cancer a contagious disease ?).

4- Find - a - Generality level (FG) in which students are required to derive or invent a new generality from new instances that students haven't seen before (e.g., derive the principle of lightning by manipulating these new phenomena?) (see, Darwazeh, 1997b, p.110-115; Hill, 1984, p.183 for more details).

Educators (e.g.; Darwazeh, 1996) believe that presenting different levels of adjunct questions during instruction will induce different levels of students' learning. Adjunct questions could be presented on different levels according to the cognitive processes they need. These levels usually vary from simple (Remember - an - Instance RI and Remember - a - Generality

RG) to complex (Use - a - Generality UG) and Find - a - Generality FG). Simple-level questions such as (RI, RG) usually require low-cognitive processes, whereas complex-level questions such as (UG) usually require high cognitive abilities. Some studies have supported this hypothesis Watts & Anderson (1971), and Shavelson et al. (1974) , for example, found that the subjects who received high-level questions in which they were asked to apply certain principles to new instances during instruction performed significantly better on the posttest than the subjects who received low-level questions requiring them to recall specific information. Rayan (1973) concurred with the above results and emphasized that asking different levels of adjunct questions will promote different levels of learning. He found in his study that high - level questions lead to a greater effect than low - level ones on the students' learning. Winne (1979), on the other hand, could not find a significant difference between the effect of high versus low - level questions in terms of their effects on students' learning.

Educators (Darwazeh, 1982; Hudgings, Dorman, Harris, 1979; Loring, 1987; etc.) also believed that the effect of adjunct questions is influenced by different levels of students' ability. Martin et al. (1994) summarized Barnes' study (1978) in which he investigated the interaction between the levels of questions - high levels (UG,FG) and low levels (RI, RG) - and the levels of student's ability (high versus low). Results showed that there was a positive relationship between higher level questions and higher level student's ability.

Darwazeh (1982) on the other hand, found a significant interaction between the position of questions and student's ability which indicated that pre-questions enhanced the learning of low ability students, whereas post-questions enhanced the learning of high ability students. Hudgings, Dorman, and Harris (1979) also had found similar results which indicated that post-

questions enhanced learning of high ability students, whereas pre-questions did not make a difference between low or high ability students.

In the current study, the researcher intended to investigate different types of questions: essay questions versus objective ones and three different levels of questions: Remember - an - Instance, Remember -a- Generality, and Use - a - Generality, on three levels of learning (RI, RG, UG) based on Merrill's taxonomies of learning. The Find - a- Generality level (FG) will not be included in this study because it has low relevance to most classroom objectives, and because "Merrill" has not yet adequately specified the nature of questions for the Find- a - Generality (FG) level (See, Darwazeh, 1982).

Different levels of student's ability (high, medium and, low) that were measured by the General Secondary Examination (Tawjihi exam) average will also be studied in the current study. Thus, the researcher will classify student's ability into three levels: high-level ability students are those whose average is 85% and more; medium -level ability students are those whose average ranged from 84 % to 75 %; and low- level ability students are those whose average is 74% and below. This classification will be based on the general average in the Tawjihi exam, as we have just mentioned.

Research problem and purposes

= = = = = = = = = = = = = = = = = =

It is known that students are always afraid of tests which assess their knowledge that they are taught in classrooms. Many students attribute their low achievement to the types of tests that the teachers use for evaluation.

Teachers are also puzzled as to what type of questions they have to use in measuring their students' learning. On the other hand, previous studies

which have investigated essay versus objective type tests did not differentiate among the different levels of learning that the posttests may measure.

Therefore, the purpose of this study is to investigate the effect of question <u>types</u> (essay versus multiple-choice), and question <u>levels</u> (Remember - an - Instance (RI), Remember - a- Generality (RG), and Use - a- Generality (UG) on three levels of learning based on Merrill's taxonomy (RI, RG, UG). Student's ability levels (high, medium, and low) will also be investigated in this study by using Three -Way Analysis of Variance (ANOVA) (See the method section).

Significance of the study

= = = = = = = = = = = = = =

The Findings of this study are expected to be of great importance for educators in general and teachers in particular who are interested in testing and evaluating students' learning. Therefore, the most significant aspects of this study are offering insights to

* the type of questions (essay or objective) which most affects students' levels of learning,

* the level of questions (RI, RG, or UG) which most affects students' levels of learning, and

* the student's ability (high, medium, or low) which most affects students' levels of learning.

Research questions

= = = = = = = = = = = =

The research questions are the following:

1- Which type of questions (essay or objective) has more effect on students' overall learning ?

2- Which level of questions (RI, RG, or UG) has more effect on students' overall learning ?

3- Which level of ability (high, medium, or low) has more effect on students' overall learning ?

4- Is there an interaction between question types (essay versus objective) and question levels (RI, RG, UG) on students' overall learning ?

5- Is there an interaction between question levels (RI, RG, UG) and student's ability (high, medium, low) on students' overall learning ?

6- Is there an interaction   between question types (essay versus objective) and student's ability on students' overall learning ?

7- Is there an interaction between types and levels of questions and student's ability on students' overall learning ?

8- On the posttest, is there a significant difference between the control groups' performance which didn't receive questions during the experiment and the experimental groups' performances which manipulated questions during the experiment ?

Research hypotheses

= = = = = = = = = = = =

The researcher postulated the following eight null hypotheses:

1. There are no significant differences at (0.05) a priori level of significance between the essay questions and the objective ones on all levels of learning (RI , RG , UG).

2. There are no significant differences at (0.05) a priori level of significance among the three levels of questions (RI, RG, UG) on all levels of learning.

3. There are no significant differences at (0.05) a priori level of significance among the effects of the three levels of the student's ability (high, medium, and low) on all levels of learning.

4. There is no significant interaction at (0.05) a priori level of significance between question types and their levels on all levels of learning .

5. There is no significant interaction at (0.05) a priori level of significance between question types and student's ability on all levels of learning.

6. There is no significant interaction at (0.05) a priori level of significance between question levels and student's ability on all levels of learning.

7. There is no significant interaction at (0.05) a priori level of significance between question types, levels, and students' ability on all levels of learning.

8. There is no significant difference between the control groups' performance compared with the experimental groups' performance.

Chapter Two
Review of Literature

= = = = = = = = = = = =

Adjunct Questions have received a great interest of educators and instructors since they are considered an effective teaching instrument, hence, an effective evaluation device (e.g., Al-Smadi 1992; Bridgeman and Lewis,

1994; Darwazeh, 1997a, 1997b; Harder, 1991; Kneip & Grossman 1979; Zeidner 1990; etc.).

Accordingly, adjunct questions, as a tool of assessing students' levels of learning, has been tackled by many specialists, educators, and researchers .

There have been a number of studies designed to compare between multiple-choice and essay questions and to examine their effects on students' achievement. In this section, it would be beneficial to summarize the results of some previous studies that have been conducted on adjunct questions in terms of their types, levels, and their interaction with student's ability.

The author of this research has reviewed these studies and classified them into three main headings: (1) type of adjunct questions, (2) level of adjunct questions, and (3) the interaction among question types, levels and students' ability.

Types of Adjunct Question Studies
= =

The question formats that are mostly used in the construction of achievement tests at both school and college levels could be classified into two main categories: (a) objective type   format   such as   multiple-choice

questions, and (b) essay type format such as short - answer questions (Darwazeh, 1997, p. 60 ; Gronlund and Linn, 1990, pp. 166-200).

Some educators said that an essay type test is a better measurement tool of learners' achievement than an objective one, because: 1) it requires the learner to produce the answer not to choose it, 2) it can measure different levels of students' learning , 3) it is relatively easier to prepare than multiple-choice type, and 4) it is the only means that we can use to assess the high levels of learning such as application, analysis, synthesis, problem solving, and evaluation (See, Ebel & Frisble 1979, p.100; Hambleton & Murphy, 1992). Other educators assume that the multiple-choice type test is more popular because it: 1) requires the learner to discriminate among the alternatives easily and choose the right answer, 2) removes subjectivity in scoring answers, and 3) reduces measurment error and elaminates random guessing by implementing a certain formula (Bridgeman & Rock 1993; Hambleton & Murphy,1992; Hopkins & Stanely, 1981, pp.231-234; Mehran & Lehman, 1978, pp. 208-300; Kubiszyn & Borich, 1990, pp. 84 - 114; etc.).

Bridgman and Lewis (1994); Darwazeh (1997a) have also differentiated between essay and multiple-choice test items; they stated that essay questions assess productive and organized skills that can't be measured by multiple-choice questions. Besides, the two types differ in their coverage of the content domain; essay questions usually cover a smaller range of learned content, while multiple-choice questions cover a broader range of learned content.

There has been a number of studies which were conducted to compare between multiple-choice and essay questions to investigate their effects on students' achievement. For example, Roderick & Anderson (1968) compared between the effect of essay versus multiple-choice questions on

students' achievement. Students were asked to read a 3400 word passage on "Classical Conditions" and to answer the questions that follow the passage. Students were distributed into two experimental groups: one received essay post questions and the other received multiple-choice post ones. Results showed that students who received post essay questions performed better than those who received multiple-choice post ones on the post test which measured the retention learning (See, Darwazeh,1997a, p.149).

Arrasmith, Sheehan, and Applebaum (1984) supported Roderick & Anderson's results when they compared between the selected - response strategy (multiple-choice) and the constructed - response strategy (essay) for assessing a third - grade writing task. The performance of 371 third - grade students were measured by selected - response test items (multiple-choice) and constructed - response test ones (essay) to assess the mastery of the written composition.Results indicated that essay items were better for mastering the writing composition than multiple-choice ones.

Foos and Fisher (1988) found the similar results when they conducted a study to assess the value of test taking as a means of enhancing rather than monitoring learning. 105 college students read a short essay about the "American Civil War" and were then given either a postpassage questions or no quetions at all on the text material. The form of the postpassage questions was either fill-in (essay) or multiple-choice; and the knowledge examined was either directly stated in the original text (verbatim) or could be logically derived (inferential) from the text. A final test, containing the above mentioned levels of learning, was given to all students two days later. Results of the final test indicated that students who received inferential (UG) fill-in (essay) questions performed better than those who received verbatim (RI, RG) multiple-choice questions.

Hambleton and Murphy (1992) found the same results as those found by Foos and Fisher when they investigated whether the multiple-choice question type has a positive effect on students' achievement and if they can measure high order cognitive skills copmared with the authentic   measurement (essay). The researchers collected data from several articles on essay and multiple-choice items to assess the effectieveness of the multiple-choice question over the essay one or vice versa. After comparing the data, they found that the authentic measurement items (essay questions) have a little effect over and above that provided by the multiple-choice items on students' achievement when assessing higher order levels.

Lukhele, Thissen, and Wainer (1994) agreed with Hambleton and Murphy's results when they investigated the relative value of multiple-choice, and constructed response (essay) items on two achievement tests in chemistry and United States History. The researchers   analyzed all of the data from the operational administration of the 1989 Advanced Placement Test in Chemistry as well as from the 1988 administration of the Advanced Placement Test in United States History. Analyses showed that the constructed response items (essay) of the tests yielded little information over and above that provided by the multiple-choice items written on a retention level of learning.

On the other hand, some researchers found that objective questions (e.g., multiple-choice) have a greater effect than essay questions on students 'learning. For example, Duchastel & Nungester (1982) conducted a study to investigate the effect of two question types (essay versus multiple-choice) on student's achievement .125 high school students in Grade 10 from a midle-class suburban high school were distributed randomly into three groups: one group received essay postpassage questions written on remember level; another

group received multiple-choice postpassage questions written on the same level; and the last group is the control group who received a true or false questionnaire concerned with general study habits. All students were asked to study a brief history text with the post questions mentioned above. On the delayed post general retention test, consisting of 24 items selected from the initial post questions, twelve items were selected from the initial short answer (essay) questions, and twelve from the initial multiple-choice questions. Thus, for students in each of the initial questions, half of the items was repeated in the same format as seen previously, while the other half was represented in the alternate format. The students were told that the test would require them to write down the main points they could remember, and that there would be specific questions of details as well. Results showed that students who received multiple-choice questions performed better than those who received essay questions and than those who were in the control group, on a post test that measured retention learning.

Perkins (1984) reached similar result as Duchastel & Nungester (1982) when he analyzed two common types of questions in testing English as a Foreign Language (EFL) reading comprehension. These tests were: 1) objective question types (true-false, multiple-choice), and essay ones (missing letters, and grammar paraphrase). The subjects were 19 Egyptian adults from the American University in Cairo, who were enrolled in an intermediate-level English as a Foreign Language. Subjects were distributed randomly into four groups: the first group received true-false posttest questions, the second group received multiple-choice posttest questions, the third group received missing letters posttest questions, and the fourth group received grammar paraphrased posttest questions. All subjects were asked to read a 232 word reading passage selected from their required text. Results

of the posttest showed that objective items produced better results than did essay ones on a comprehension level of learning.

Bridgman and Lewis (1994) agreed with Perkins' results when they investigated the relationship between essay and multiple-choice questions. 32 college students from public and private institutions, who entered the college in the fall of 1985, were tested in four subjects: American History, European History, English Language and Composition, and Biology by using two types of questions: essay questions and multiple-choice ones. Results indicated that multiple-choice scores of the American History and Biology examinations were superior to the essay ones, but essay scores were essentially equivalent to multiple-choice scores in European History, and English Language.

Some researchers didn't find significant differences between essay questions and objective ones on students' learning. For example, Frase (1968); Williams (1963) could not find significant differences between the effect of short answer (essay) questions and the effect of multiple-choice ones on the student achievement (See , Darwazeh , 1987, p.111, for more details).

Brigdman (1992) did not also find a significant difference when he compared between open-ended (essay) and   multiple-choice formats. In a separable data collection, 364 paid volunteers who had taken the Graduate Record Examination used a computer keyboard to enter answers to the same set of questions. Results indicated that the total scores for the multiple-choice and open-tests (essay) demonstrated remarkably similar effects on achievement.In other words, there were no significant differences between the effect of question types (essay versus multiple-choice) on students' learning on retention type test.

Bridgeman and Rock (1993) reinforced Brigdeman's findings when they conducted a study to explore the relationship between multiple-choice and open ended (essay) analytical questions. The scores of 349 students of the Graduate Record Examination General Test were analyzed. Results showed that the open ended (essay) items were not measuring anything beyond what is measured by the multiple-choice version of these items which means that the two question types (essay versus multiple-choice) have the similar effect on the student's achievement .

Thissen, Wainner, and Wang (1994) agreed with Bridgeman and Rock's results when they didn't find a significant difference between multiple-choice and essay questions on the student achievement. They used 2000 students who took Computer Science and Chemistry tests of the College Board's Advanced Placement Program, and divided them into two groups: one group received multiple-choice questions, and the other one received essay questions. Results showed that essay sections have the same effect on students' achievement as the multiple-choice on solving problem (application level) test.

Summary

= = = = =

From the above studies, the researcher concluded that some researchers (e.g., Arrasmith, Sheehan & Applebaum, 1984; Hambleton and Murphy, 1992; Lukhele, Thissen, and Wainer, 1994; and Roderick & Anderson, 1968) found that essay type questions have more effect on students' learning than do multiple-choice ones especially on high levels of learning (UG). Others, (e.g., Bridgman & Lewis, 1994; Duchastel & Nungester, 1982; and Perkins, 1984; etc.) found that multiple-choice type questions have more effect than essay

ones especially on remember levels of learning (RI , RG). But some researchers found that there were no diferrences between the effect of essay versus multiple-choice question types on all levels of learning (e.g., Bridgman, 1992; Bridgman & Rock, 1993; Frase, 1968; Thissen, Wainner & Wang , 1994; Williams, 1963).

Levels of Adjunct Quetion Studies

= =

Adjunct questions are usually designed to measure the variation of different cognitive levels. Thus, they vary from simple level such as Remember an Instance (RI) and Remember a Generality (RG) to complex level such as Use a Generality (UG) and Find a Generality (FG) according to the cognitive process the learner employs during answering the required question (Merrill ,1983). Low- level questions like remember - an - instance and remember - a - generality have been consistently used in school and teachers' textbooks, but high level questions like apply and find a generality levels have seldom been used. Thus, adjunct questions which require recall of specific information or facts (RI,RG) may produce a different level of learning from adjunct questions which require students to apply (UG) or transfer the learned idea to new situations (FG). Simply, low levels of adjunct questions induce low levels of learning and high levels of adjunct questions induce high levels of learning (Darwazeh, 1992, 1995; Martin, 1979).

There has been a number of studies which were conducted to investigate the levels of adjunct questions which teachers use in evaluating their students' achievements. In the measurement of reading students' achievement, Hoeppel (1981) conducted a study aiming at categorizing questions found in reading skills development books used in Maryland Community College Development Remedial Programs via the " Bloom Taxonomy of Educational Objectives Domain." A random sample of (555) questions was selected for the analysis.

The questions were classified according to Bloom's Taxonomy. The classification showed that 145 questions (26%) were for knowledge, 408 questions (74%) were for comprehension, 2 questions (0.0035%) were for application and no questions were for analysis, synthesis and evaluation. In other words, results showed that (99%) of the questions were categorized within two levels: knowledge and comprehension.

Al-Makzoomy (1986) agreed with Hoeppel's results when he analyzed the Jordanian secondary- school teachers' responses through a questionnaire, on the levels of questions used in teaching reading comprehension. Results showed that (68%) of the teachers usually place more emphasis on literal-type questions (remembrance) than on inferential-type questions (application).

Rinser (1987) supported the results of the above two studies (Hoeppel,1981; Al-Makzoomy,1986) when he studied the cognitive levels of questions demonstrated by test items that accompanied fifth grade science textbooks. Based on Bloom's taxonomy, results showed that about (95%) of the test questions devoted to knowledge or comprehension (RI, RG), but (5%) were used for application (UG), and (0.2%) for evaluation, analysis, and synthesis questions were neglected completely (p.126).

Tollefson (1989) summarized the results of some studies which investigated the relationship between teachers' questions and students' learning. These results were:
1. The cognitive complexity of the students' responses was largely determined by the cognitive complexity of the questions.
 2. Recall and recognition (RI, RG) questions, which have a low level of cognitive processing, generally require simpler responses than higher level

(UG) questions requiring students to express opinions or to provide interpretations (pp. 6-7).

Harder (1991) agreed with "Al-Makhzoomy's" results when he studied the levels of classroom oral questions which were used by the teachers of Arabic language at the basic educational stage in Jordan. A sample of 60 teachers were given 45 minutes to ask their students different levels of questions. Results, which were classified according to Bloom taxonomy, showed that the percentage of questions at the level of knowledge was   47.5% of the total questions, 32.4% for the comprehension questions, 13.1% for application questions, 5.6% for analytic questions, 1% for synthetic questions, and finally, .7% for evaluation questions.

Perry, VanderStoep, and Yu (1993) conducted a study to investigate the levels of questions that were asked in first grade addition and subtraction lessons in Japan, Taiwan, and the United States. Some researches have argued that knowledge is, in part, constructed through questions and these may be used differently in U.S. than in Asian classrooms. Thus, each question of addition or subtraction in 311 observed lessons was coded as 1 of 6 levels of questions. Analyses revealed that the Asian teachers asked significantly more questions about conceptual knowledge than did U.S. teachers. In addition, Chinese teachers asked significantly more questions that were embedded in a concrete context (verbatim) than did U.S. teachers .

Martin, et al. (1994) reinforced Harder's (1991) results. Martin and her friends stated that teachers usually use questions which require factual answers and low levels of thinking (i. e., knowledge and comprehension questions) make up at least 70% of the questions, while questions that require application, analysis, synthesis, or evaluation thinking are used much less often (p.156).

Watts and Anderson (1971) investigated the effect of questions that require students to apply (UG) what they have read to new situations. 300 high school seniors answered an inserted post questions after reading a passage that consisted of a 450-word passage explaining psychological principles. Subjects were distributed randomly into three groups: one group received application questions (UG), another group received name questions (RI, RG), and the control group was asked to read the passage only. Results showed that subjects who received questions requiring them to apply (UG) principles to new examples performed better on the posttest (which measured application questions) than did subjects who received questions requiring them to recall (remember) previous examples.

Rickards & Vesta (1974) found similar results to Watts and Anderson's results when they studied the effect of question levels on learning. The subjects were 80 college sophomores enrolled in an introductory educational psychology course. Each subject responded to one of the four types of questions: rote - learning - of - facts, rote - learning - of- ideas (RI, RG), meaningful - learning , or task - irrelevant questions (UG). Results showed that students who received questions on meaningful learning (UG), which required subsumption or organization of facts under given ideas, performed better than students who received questions requiring them to recall specific facts or ideas (RI or RG) - rote learning of facts and ideas - on a posttest containing these levels of learning .

Kneip & Grossman (1979) summarized some studies which reported that students' achievement was significantly and positively affected when teachers use mostly high level questions. For example, Rayan (1973) compared between the effects of high and low level of questions on the social

studies achievement of fifth and sixth grade students. Results indicated that questions which demand high cognitive levels (UG), beyond the recall level, were superior to the low level (RI, RG) questions in producing not only the high level understanding but also producing high levels of achievement.

Redfield (1981)conducted a study to examine the effect of teachers' questions on student achievement. In this study, twenty studies on teachers' use of higher (UG) and lower (RI, RG) cognitive questions were reviewed . Higher cognitive questions require the student to manipulate information to create and supply a response; lower cognitive (RI, RG) questions call for verbatim recall or recognition of factual information. Results of the studies reviewed showed that teachers' use of higher cognitive (UG) questions had a positive effect on student general achievement on a retention level of learning.

Royer and Konold (1984) examined Hunkin's (1969) study in which he investigated the effect of two levels of questions, knowledge (low - level) and evaluative (high - level) on students' achievement in two groups. One group studied social materials provided with knowledge questions, the other group studied the same passage but was provided with evaluative questions. After four weeks, the whole students took an examination consisting of questions of all six levels of Bloom's Taxonomy. Results showed that the two groups did not differ on items from the lower taxonomic levels, but they differ on evaluative items. That is, students receiving higher level evaluation during the study phase (UG) performed significantly better on high level questions in the posttest .

Al-Nayef (1989) supported Hunkin's results when he conducted a study to investigate the effect of adjunct question levels on reading comprehension of the eleventh graders. Al-Nayef used two levels of questions: low level

questions (knowledge), and high level ones (i.e., comprehension, application, analysis, synthesis, and evaluation). The sample of his study consisted of 144 eleventh grade scientific students in Al-Koureh District Education Directorate in Jordan for the second semester of the academic year 1988 /89. Subjects were randomly assigned to three groups: High level experimental group (N = 82), low level experimental group (N =35), and control group (N = 27). Results showed that students who were exposed to high level questions (comprehension, analysis, synthesis, evaluation) performed better on the same levels of questions in the post test.

On the other hand, some researchers found that low level questions have a great effect on students' achievement. For example, Felker and Dapra (1975) have investigated the effect of different levels of adjunct questions on students' learning. 93 introductory psychology students at the Greensburg campus of the University of Pittsburgh were distributed randomly into three groups: one group received verbatim questions (RI , RG), another group received application level questions (UG), and the third group was the control group which received no adjunct questions during instruction but were asked to read the passage only. Results showed that subjects who received comprehension post-questions requiring them to recall (remember) the text performed significantly better than those who received application - level (UG) adjunct questions requiring them to identify new examples of learned concepts or principles on application level posttest, and better than the control group which received no adjunct questions during instruction.

Samson et al. (1987) conducted an analytical study to investigate the result of Redfield & Rousseau's (1981) study which showed that there was a large effect of higher order (UG)questions on student achievement. The main aim of Samson et al.(1987) study was to test the effects of question levels on

student achievement and to examine Redfield & Rousseau's results. They analyzed fourteen studies which were conducted to examine the effects of different adjunct question levels on learning. The results of the fourteen studies were examined and compared with the results of Redfield & Rousseau. Results of Samson's study indicated that higher level questions (UG) have a small positive effect on learning. This result contradicts the previous results which indicate the opposite.

Perkins et al. (1990) agreed with Samson's analysis when they investigated the effect of question levels on English as a Second Language (ESL) reading comprehension. A sample of 150 Japanese English-as-a-second- language students at Southern Illiois University, Carbondale was given a reading comprehension test containing three levels of questions: factual (RI), generalization (RG), and inference (UG), to measure comprehension effects at different proficiency strata . Results showed that there were significant differences for the factual questions (RI), but no significant differences among the generalization (RG), and inference (UG) levels of questions.

Some researchers did not find a significant main effect with respect to the effect of adjunct question levels on students' achievement .For example, Andre et al. (1980) have done seven experiments to explore the effect of application (UG) question levels and factual ones (RI , RG) on students academic achievement. The subjects were (120, 155, 87, 57, 135, 194, 194) male and female students attending Ames High School in Ames, Iowa; they were asked to read a short passage of three concepts and answer two levels of questions: remembrance (RI, RG) and application (UG). Results showed that five studies found no significant differences between question levels

(remembrance and application). Only two studies found that remember - levels (RI , RG) adjunct questions were superior to application (UG) ones.

Darwazeh & Reigeluth (1982a) have also studied the level and position of adjunct questions and their effects on memory and application learning. Seventy- four male and female eighth grade students, at a suburban school in Syracuse, New York, were presented with a 450 - word passage from a social studies textbook. Nine short - answer posttest questions were written on three levels of learning (RI, RG, and UG), and were used either before or after the relevant passage. On the nine - short answer posttest questions, which were comprised of these three levels of learning (RI, RG, and UG), they found that there were no significant main effects with respect to the question levels.

Darwazeh (1982b) didn't also find any significant difference among the performance of the experimental groups who received either remember - an-instance (RI), remember a generality (RG) questions and the performance of the experimental ones who received use a generality questions (UG), though the direction of the groups' mean indicated that RI level had higher (x= .48) effect on students' performance than RG (x= .41) and UG (x= .39).

Summary

======

From the above revision of previous studies, we can conclude that some researchers found that high level questions (i. e., use a generality) have a greater effect on students' achievement than low level ones (e.g., Al-Nayef, 1989; Kneip & Grossman, 1979; Redfield, 1981; Rickards & Vesta, 1974; Royer & Konold, 1984; Watts & Anderson, 1971). Others found that low level questions have a greater effect on students' achievement than high level ones (e.g., Felker & Dapra, 1975; Perkins et al.,1990; Samson et al., 1987).

Few studies did not find significant differences among the different levels of questions: remember an instance, remember a generality, or use a generality on later learning (e.g., Andre et al. 1980; Darwazeh & Reigluth 1982a; Darwazeh, 1982b).

Interaction between types and levels of adjunct questions

= =

Researchers and educators have postulated that using the two types of different questions (essay versus multiple-choice) will lead to different levels of learning outcomes, and different levels of questions (remember, comprehension, application, etc.).

The depth of cognitive processing was influenced by the type of questions on the one hand and the level of questions on the other hand .

Some educators claim that multiple-choice questions usually induce low levels of learning such as RI and RG, whereas, the essay questions usually induce high levels of learning such as UG and FG (Hancock, 1994).

Anderson & Biddle (1975) expected that essay questions could improve students' achievement if they are written on high levels of learning (use a generality and/or find a generality), whereas, multiple-choice questions can improve students' achievement if they are written on low levels of learning i.e., remember an instance and/or remember a generality levels.

Darwazeh (1987, 1997) stated that essay questions have greater effect on students' achievement than multiple-choice ones especially when they are written on high levels such as evaluation, synthesis, solving problems, etc. (p. 79).

Al-Smadi (1992) indicated that the use of question type (essay or multiple-choice) of low - cognitive levels (RI, RG) does not help students learn on high levels of learning because these levels of questions generally elicit short responses and require no deep thinking .In other words, they don't require students to comprehend, interpret, or use the information in new situations. Whereas, the use of question type (essay or multiple-choice) of high cognitive levels (UG) plays an important role in helping students to understand the text because they generally elicit long complex, and deep responses; they require students to comprehend, interpret, and use the information properly (pp.12-13).

Hopkins & Stanely (1981) summarized the results of some studies (e.g. Hoffman, 1962; Lafave, 1966) which showed that objective tests (multiple - choice) can measure mostly knowledge of facts, whereas essay tests can measure more complex, higher levels of understanding and have a great effect on learning (See, Hopkins & Stanely, 1981, p. 202).

Concerning the interaction among question levels and students' ability, Sanders (1973) found significant interactions which indicated that low ability students performed better on remember questions;   whereas, high ability students performed better on application level. Sanders used 72 college students, and 40 item multiple-choice immediate and delayed test measuring relevant and irrelevant information (See, Darwazeh,1996, for more details). Memory (1983) supported the above result. He used a posttest containing 29 questions written on literal (remember) and application levels. Results indicated that there was a significant interaction which indicated that high ability students performed better on the post test when they are written on high levels (application).

Darwazeh (1982) also reached similar results, when she investigated the interaction between students' ability and the position of adjunct questions. She used 181 tenth-grade students, and a 19-item short-answer (essay) questions measured remember (RI, RG) and application levels of learning. Results showed that high ability students performed better on high level (UG) post-passage questions, while, low ability students performed better on low level pre-passage questions.

Summary

= = = = =

Although there were few studies which addressed the interaction between the type of questions (essay versus multiple-choice) and the level of questions (RI, RG, and UG) and its interact with students' ability, we can conclude from the above studies (e.g., Darwazeh, 1992; Hancock, 1994; Hopkins & Stanely, 1981) that essay questions have great effects on students' achievement when they were written on high cognitive levels (UG). On the other hand, some researchers (Anderson & Biddle, 1974; Foos & Fisher, 1988) indicated that multiple-choice questions have great effect on students' achievement when they are written on low levels of learning. Researchers (e.g., Darwazeh,1982; Memory 1983; Sanders, 1973) found that high ability students performed better when they received essay high level postquestions; whereas, low ability students performed better when they received low level prequestions.

General Conclusion of the Research Studies

= =

From the above studies, the auther found that:

1. Some researchers indicated that essay type questions have more effect on students' learning than do multiple-choice ones especially on higher level of learning (e.g., Arrasmith, Sheehan & Applebaum, 1984; Hambleton & Murphy, 1992; Lukhele, Thissen, and Wainer, 1994; Rodrick & Anderson; etc). Others, indicated that multiple-choice type questions have more effect on students' achievement than essay ones especially on lower level of learning (e.g., Bridgman & Lewis,1994; Duchastel & Nungester, 1982; Perkins, 1984; etc.).

2. Some researchers found that higher level questions (e.g., use a generality) have greater effect on students' learning than lower level questions (e.g., Al-Nayef, 1989; Kneip & Grossman, 1979; Redfield, 1981; Rickards & Vesta, 1974; Watts & Anderson, 1971). Few researchers found that lower level questions have greater effect on students' achievement than higher level ones (e.g., Felkar & Dapra, 1975; Perkins et al.,1990; Samson et al., 1987).

3. Some researchers (e.g., Anderson & Biddle, 1974; Darwazeh, 1992; Hancock, 1994; Hopkins & Stanely,1981) indicated that essay questions have greater effect on student achievement when they are written on high levels of learning (e.g., application, evaluation etc.), whereas few researchers (e.g., Anderson & Biddle, 1974; Foos & Fisher, 1988; Haffman, 1962; Lafave, 1966) indicated that multiple-choice questions have greater effect on student's achievement when they are written on remember levels of learning (remember instance and remember generality).

After analyzing the previous studies on adjunct questions and their effects on students' achievement, the researcher did not find any study that addressed the effect of question types (essay versus multiple-choice) and levels based on "Merrill's" Taxonomy of learning (RI, RG, and UG) and their

interaction with students' ability except Darwazeh's studies (1982, 1984). Besides, the researcher believes that the present study will shed light on how to use adjunct questions during instruction in terms of their types, levels, and their interaction with students' ability levels. Since this topic hasn't yet been covered by interested researchers, the researcher intended to conduct this study to see whether type and level of adjunct questions have an effect on students' academic achievement, with a consideration of students' ability.

Chapter Three

Methodology

Chapter Three

Methodology

= = = = = = = =

This chapter consists of: 1) population, 2) sample, 3) experimental design, 4) statistical design, 5) instructional task, 6) treatments, 7) measures, 8) procedures, and 9) scoring .

Population

= = = = = = =

The   population   of this study consisted of all students who   were enrolled in "University English I" course (10103), the first semester 1996 / 1997   at An-Najah National University, Nablus, Palestine. The   whole population was (935) students, most of them were at a freshman level.

Sample

========

A random sample for this study   was   taken   from the whole population; it consisted of (140) students who came from different colleges at An-Najah National University, Nablus. Almost one third of the sample was from the Faculty of Engineering, the rest of it came from different faculties at the An-Najah. Students ages ranged from 19- 21 years, with an average of 20. See, Table (1) .

Table (1) shows the number of students who participated in the study and the name of the college which students are affiliated to.

Name of the college	Number of students
Faculty of Arts	26
Faculty of Science	13
Faculty of Economics	32
Faculty of Education	7
Faculty of Engineering	55
Faculty of Islamic Law	2
Faculty of Fine Arts	1
Faculty of Pharmacy	2
Faculty of Law	2
TOTAL	140

Experimental Design
= = = = = = = = = = = = = =

The experimental design was the posttest - Only Control Group Design. The diagram for this design appears in the following figure:

R X O
R   O

= =

R = Random assignment

X = Independent variables

O = Observation or test

Statistical Design

= = = = = = = = = = =

A Three - Way Analysis of Variance (ANOVA) factorial design (2 x 3x3) was used for analyzing the data of the study by using general F-test. The first factor was the type of questions (essay versus multiple-choice questions), the second factor was the level of questions (RI, RG, and UG), and the third factor was students' abilities (high, medium, and low). See, Table (2).

Table(2) shows the three variables of the study: question types, question levels, and student's ability.

Question types	Multiple-choice			Essay		
Question levels	RI	RG	UG	RI	RG	UG
Student's ability	High	Medium	Low	High	Medium	Low

One Way Analysis of Variance will also be used (1x7) to compare the experimental groups with the control group (See, Table 3).

Table (3) shows the experimental groups and the control group which were used in the study.

RIM	RGM	UGM	RIE	RGE	UGE	Control

RIM = Remember- an - Instance Multiple-choice.

RGM = Remember - a - Generality Multiple-choice.

UGM = Use - a - Generality Multiple-choice.

RIE = Remember - an - Instance Essay.

RGE = Remember - a - Generality Essay.

UGE = Use - a - Generality Essay.

If the general 'F' shows significance at (.05) priori level of significance, then a Post - hoc Analysis of Variance will be used by using a Scheffe test at (.05) level of significance.

Instructional Task
= = = = = = = = = =

The learning passage employed in this study was selected from the regular introductory course textbook "University English I". A - 913 - word passage, entitled "The Common Cold" contained 13 paragraphs discussed the idea of the common cold's spread and ways of its prevention. Students who participated in this study haven't had a previous knowledge of the passage. The passage mostly consisted of general information such as concepts and procedures, and specific information such as dates, names, and examples of the general information. The entire passage is presented in Appendix A (See, Appendix A).

The Experimental and Control Groups
= =

The sample of this study was divided randomly into six experimental groups and one control group. Each group consisted of (20) students. Three of the experimental groups were directed to read the passage once then answer a group of multiple-choice questions written on three different levels   of learning   based   on   Merrill's Taxonomy (RI, RG, UG). One group was directed to read the passage once, then to answer (7) multiple-choice questions of Remember - an - Instance level (RI) which required students to recognize specific information such as: Colds are caused by a: a) Germ, b) Bacterium, c) ýFangus, d) Virus. The second group was directed to read the passage once, then to   answer (6) multiple-choice questions of Remember -a - Generality level (RG) which required students to remember general information such as: A contagious disease is a disease which a)doesn't have treatment, b) spreads by touch, c) affects only pregnant women, d) causes death. The third experimental group was directed to read the passage once, then to answer (6) multiple-choice questions of Use - a - Generality level (UG) which required students   to apply a generality to a new instance such as: Medicine can be given to patients in different compositions such as injection. Another common way of taking medicine is a) drinking water, b) tasting sugar, c) tablets, d) smelling air.

The other three experimental groups were directed to read the same passage and answer the same levels of questions which the above mentioned experimental groups have read and answered, but the questions were given to them in an   essay   type. The only difference between two of them was that the first three experimental groups received questions written in a multiple-choice form, whereas the second three experimental groups received the same questions but they were written in an essay form, that is, one group read the passage and answered (7) essay questions on Remember - an - Instance (RI)

level, such as: What causes cold?). Another group read the passage and answered (6) essay questions on Remember - a - Generality level (RG), such as: What is a contagious disease?). The third experimental group read the passage and answered (6) essay questions written on Use - a - Generality level (UG), such as: Medicine can be given to patients in different compositions like injection. Name another way of taking medicine that is commonly used by people.

The control group was directed to read the same passage once without receiving any type of questions. Each booklet was given a number to keep the track of the experimental and control groups when they took the posttest.

Measures
=======

Based on Merrill's taxonomy of learning (1983) a - 19 - item posttest was written on three levels of learning (Remember- an - Instance RI, Remember - a - Generality RG, and Use - a - Generality UG), asking about the general and specific information which mentioned in the passage (See, Darwazeh (1995a) to learn more about Merrill's Theory). 10 of the 19 items were written in an essay form and 9 items, the rest of the posttest, were written in an objective form (See, Table 4).

The remember - level test items (RI and RG) were those which students have manipulated during the experiment, whereas the application level test items (UG) were different from those questions which students have manipulated during the experiment, otherwise they will be measuring the remembrance of application rather than the application level itself. The test

tried to measure all the specific and general information that came in the passage.

Table (4) shows the number of question types and levels which were used in the achievement test.

Question types	Question levels			Total
	RI	RG	UG	
Multiple-choice	3	3	3	9
Essay	4	3	3	10
TOTAL	7	6	6	19

Test Reliability
= = = = = = = = = =

The reliability of the test was computed by using "Kuder Rirhardson 21".

$$R = 1 - \frac{x - \{k - x\}}{k\ (S.D.)}$$ where : x = mean

S .D.= standard deviation

k = number of items on post- test
(See, Shohamy,1985 p.21)

In order to use this formula, we computed the following:

X= students' scores = 1822.5

X = 1822.5   140 = 13

$(x - x)$ = 4716.5

4716.5   140 = 33.7

S . D. =   33.7 = 5.8

Thus, the reliability of the test is computed by the following steps

$$1 - \frac{13 - (19 - 13)}{19(5.8)} =$$

$$= 1 - \frac{78}{639.16} = .12 =$$

$$R = 1 - .12 = \mathbf{.88}$$

Procedure
=======

This study was conducted in the second semester of the academic year 1996/97, at An-Najah National University, Nablus, Palestine. The following procedure was followed:

1. The researcher selected the instructional passage, which is entitled "The Common Cold" from the required textbook University English I , at An-Najah University, after he coordinated with the teachers who teach the course . Students haven't had a previous knowledge of the passage; they were about to take it, according to their lesson plan, within the coming ten days from the date of the experiment.

2. The researcher took the random sample of his study which consisted of (140) students and assigned them randomly into 7 groups: 6 experimental groups and one control group, each consists of 20 students.

3. For the six experimental groups, the researcher designed six booklets each consists of six to seven questions written in either an essay type or a multiple-choice type covering all specific and general information that came in the passage, and measuring three levels of learning based on Merrill's

taxonomy: Remember - an - Instance, Remember - a - Generality, and Use a Generality. For the control group, the same passage was only used with no post-passage questions (See, Table 5).

Table (5) shows the seven booklets which were designed for each of the experimental and control groups.

No. of the experimental groups which were used in the study	Type & level of questions used	No. of questions used in each group
Experimental group no.1	RIM	7
Experimental group no.2	RGM	6
Experimental group no.3	UGM	6
Experimental group no.4	RIE	7
Experimental group no.5	RGE	6
Experimental group no.6	UGE	6
Experimental group no.7 (control group)	___	—

The seven booklets were shuffled together randomly before the experiment started in order to assure the random assignment of the sample of the study.

4. The researcher had also composed an achievement test to measure the students' learning. The test consisted of 19 items that covered all specific and general information and that were written on three levels of learning based on

Merrill's taxonomy (RI, RG, and UG). These items were the same as the post-passage questions, except the Use - a - Generality questions which were different in order to avoid measuring the remember of application rather than the application level itself.

Table (6) shows the number of question types and levels which were used in the achievement test.

Question types	Question levels			
	RI	RG	UG	Total
Multiple-choice	3	3	3	9
Essay	4	3	3	10
TOTAL	7	6	6	19

5. For the suitable time needed for the experiment, the researcher consulted the English teachers who teach the required course "University English I" to estimate the length of time that students at freshmen levels need to read the passage and answer the post questions, then the posttest. The researcher also conducted a pilot study to ensure the amount of time required by freshmen students. The time of the experiment was limited as shown in Table (7).

Table (7) shows the time distribution during the experiment .

Procedure	Time allowed in minutes
Giving directions	5
Papers distributions	5
Reading the passage only	25
answering post-questions	10
collecting the booklets	5
test distribution	5
answering the test	30
test collection	5
Total Time	90 minutes

The experiment took place during four consecutive session; 90 minutes for each, starting from 8:00 a.m. until 2:00 p.m., in one day.

The researcher, who himself conducted the study, mentioned to the examinees the purpose of his study which was to investigate the effect of question types and levels on students' learning in order to enhance students' learning. The researcher also instructed the students to read and follow the instructions which were written on the front page of each booklet. Then, the researcher distributed the booklets within 5 minutes. Students were given 25 minutes to read the passage for once, and 10 minutes to answer the post-passage questions. The control group students were asked to read the passage once within 25 minutes and to wait for the test paper for 10 minutes doing something else like reading in other textbooks.

After the first 25 minutes, the researcher collected all the booklets within 5 minutes. Then, the researcher distributed the test within 5 minutes and reminded the students to read the general instructions which were written on

the front page of the test. Students were instructed several times to write the numbers of the booklets on the top left side of their tests' papers which they have just manipulated. Students were given 30 minutes to answer the posttest. After students finished answering the test, the researcher collected the papers test from all students at once. The same procedures were followed in the next three classes. Most students finished the experiment within 90 minutes.

Scoring

=====

The researcher himself graded the post test blindly as students were not asked to write their names. A sample of the test forms was also re-scored by another researcher in order to establish the reliability of the scoring procedures. The scores of the test were as shown in the following table (8). The scores of each question item were distributed with the help of the study supervisor (See, Table 8).

Table (8) shows the scores which were given to each sub-test questions.

Sub-Test Questions	Scores
Remember an Instance Essay Sub-Test (RIE)	4

Remember an Instance Multiple-choice Sub-Test (RIM)	3
Remember a Generality Essay Sub-Test (RGE)	5
Remember a Generality Multiple-choice Sub-Test (RGM)	3
Use a Generality Essay Sub-Test (UGE)	4
Use a Generality Multiple-choice Sub-Test (UGM)	3
TOTAL	22

Chapter Four

The Results

Chapter Four

The Results

= = = == = =

The data of this study were analyzed by using Three Way Analysis of Variance (2 x2 x3) to investigate the effect of question type (essay versus objective), question level (RI, RG, UG) and student's ability levels (high, medium, and low). One Way Analysis of Variance (1 x 7) was also used to compare between the control and the six experimental groups.

The analysis of variance was conducted independently on each sub-test which measured each level of learning: remember an instance, remember a generality and use a generality, and on the total test which included all sub-tests for both types of tests multiple-choice and essay tests.

Results of the Remember - an -Instance Multiple-Choice Sub-test (RIM)

On the Remember - an-Instance Multiple-Choice Sub-test (RIM), general 'F' revealed a significant effect for only question levels (P< .0002) and student's ability levels (P < .003) (See, Table 1) .

Table (1). A Three-Way ANOVA summary table for question types, levels, and student's ability on RIM sub-test.

Source of models	DF	SS	MS	F-test	P value
Question type (A)	1	0.055	0.005	0.184	0.6687
Question level (B)	2	5.391	2.695	9.079	0.0002 **
AB	2	0.143	0.072	0.242	0.7858
Student's ability (C)	2	3.656	1.828	6.158	0.0030 **
AC	2	0.004	0.002	0.007	0.9926

BC	4	1.644	0.411	1.385	0.2445
ABC	4	0.997	0.249	0.85	0.5031
Error	102	30.28	.297		

The post hoc ANOVA by using a Scheffe test showed at .05 level of significance that the mean of the experimental group who received questions on Remember a Generality (RG) level (x= 2.95) was higher than the mean of those who received either Remember an Instance (RI) questions   (x =2.62), or Use a Generality ones (x =2.47). The last two means did not differ significantly (See, Table 2).

Table (2). Means, standard deviations and number of students in each cell for question types and question levels on a RIM sub-test.

Q. types		Question levels			
		RI	RG	UG	TOTAL
Multiple	x	2.6	3	2.5	2.7
	(SD)	(.82)	(0)	(.76)	(.53)
	n	20	20	20	60
Essay	x	2.6	2.9	2.4	2.6
	(SD)	(.49)	(.31)	(.61)	(.47)
	n	20	20	20	60
TOTAL	x	2.6	2.9	2.4	7.9
	(SD)	(6.5)	(.15)	(.65)	(2.4)
	n	40	40	40	120

* P > .05
** P > .01

In terms of student's ability levels, Scheffe test showed that the mean of high ability students was higher significantly than the mean of low ability students (x = 2.82 versus x=2.28), and the mean of medium ability was higher than the mean of low ones (x = 2.68 versus 2.28). Whereas, Scheffe test did not reveal a significant difference between the mean of medium ability and the high one (x = 2.68 versus 2.82) (See, Table 3).

Table (3). Means, standard deviations, and number of students for question types, and student's ability on a RIM sub-test.

	Student's ability			
Question Type	**high**	**medium**	**low**	**TOTAL**
Multiple-choice x	2.86	2.66	2.3	2.7
(SD)	(.425)	(.65)	(.717)	(.597)
n	29	21	10	60
Essay x	2.97	2.7	2.27	2.66
(SD)	(1.22)	(1.3)	(.944)	(1.15)
n	29	20	11	60
TOTAL x	2.82	2.68	2.28	2.68
(SD)	(.822)	(.975)	(.83)	(.87)
n	58	41	21	120

Results of the Remember - a - Generality Multiple-Choice Sub -test (RGM)

On the Remember - a - Generality Multiple-Choice Sub-test (RGM), general 'F' revealed significant differences for question type (P< .0445) and student's ability (P<.0008); 'F' test also revealed a significant interaction between question types and question levels (P< .0025) which indicated that groups who received remember instance essay questions performed higher than those who received remember instance multiple-choice questions, whereas the performance did not differ for those who received either multiple-choice or essay questions measuring RG or UG levels (See, Table 4 & 5).

Table (4). A Three-Way ANOVA summary table for question types, levels, and student's ability on a RGM sub-test.

Source	DF	SS	MS	F-test	P value
Question type (A)	1	2.37	2.37	4.14	0.0445 *
Question level (B)	2	2.03	1.01	1.77	0.1753
AB	2	7.30	3.65	6.36	0.0025 **

Student's ability (C)	2	8.77	4.38	7.64	0.0008 **
AC	2	2.24	1.12	1.95	0.1468
BC	4	4.22	1.05	1.84	0.1266
ABC	4	1.12	0.28	0.49	0.742
Error	102	58.48	.573		

The post hoc ANOVA (by using a Scheffe test at (.05) level of significance) did not show significant differences for question type: essay (x= 2.23) and multiple-choice one (x = 2.43), though the essay type mean was higher than the multiple-choice one (see, Table 5).

Table (5). Means, standard deviations, and number of students in each cell for question types, and levels on a RGM sub-test.

Q. types		Question levels			
		RI	RG	UG	TOTAL
Multiple	**x**	1.8	2.4	2.5	2.2
	(SD)	(1)	(.75)	(.60)	(.78)
	n	20	20	20	60
Essay	**x**	2.6	2.4	2.3	2.4
	(SD)	(.75)	(.75)	(.91)	(.80)
	n	20	20	20	60
TOTAL	**x**	2.2	2.4	2.4	2.3
	(SD)	(.87)	(.75)	(.75)	(.79)
	n	40	40	40	120

On the other hand, Scheffe test showed significant differences at (.05) for student's ability which indicated that the high student's ability mean was higher than the low student's ability one (x = 2.55 versus 1.85), whereas, Scheffe test didn't reveal any significant difference neither between low and medium student's ability means (x = 1.85 versus 2.26) nor between high and mid student's ability means (x = 2.55 versus 2.26). See, Table (6).

Table(6). Means and the number of students in each cell for question types, and students' ability on a RGM sub-test.

Question Type		Student's ability			
		high	medium	low	TOTAL
Multiple-choice	x	2.58	2.04	1.6	2.23
	n	29	21	10	60
Essay	x	2.51	2.5	2.09	2.43
	n	29	20	11	60
TOTAL	x	2.55	2.26	1.85	2.33
	(SD)	(.68)	(.83)	(1)	(.83)
	n	58	41	21	120

Results of the Use - a - Generality Multiple-Choice Sub-Test (UGM)

On the Use - a - Generality Multiple-Choice Sub-Test (UGM), general 'F' revealed a significant difference for student's ability only (P < .0007) (See, Table 7).

Table(7). A Three-Way ANOVA summary table for question types, levels, and student's ability on a UGM sub-test.

Source of models	DF:	SS	MS	F-test	P value
Question type(A)	1	0.559	0.559	0.845	0.360
Question level(B)	2	0.874	0.437	0.660	0.519
AB	2	0.139	0.069	0.105	0.900
Student's ability (c)	2	10.34	5.170	7.810	0.0007 *
AC	2	0.867	0.434	0.655	0.521
BC	4	1.450	0.362	0.547	0.701
ABC	4	5.235	1.309	1.977	0.103
Error	102	67.52	.66		

The post hoc ANOVA by using Scheffe test showed at .05 level of significance that the mean of high ability students was higher than the low one (x = 2.37 versus 1.57), whereas, there were no significant differences neither between low and medium student's ability means (x = 1.57 versus 2.07) nor between high and medium student's ability means (x = 2.37 versus 2.07) (See, Table 8).

Table (8). Means and number of students in each cell for question types, and student's ability on a UGM sub-test.

Question types		high	medium	low	TOTAL
		Student's ability			
Multiple-choice	x	2.31	1.85	1.6	2.03
	n	29	21	10	60
Essay	x	2.44	2.3	1.54	2.23
	n	29	20	11	60
TOTAL	x	2.37	2.07	1.57	2.13
	(SD)	(.72)	(.93)	(.81)	(.82)
	n	58	41	21	120

Results of the Remember - an - Instance Essay Sub-Test (RIE)

On the remember - an - instance essay sub-test (RIE), general 'F' revealed significant differences for question types (P< .0006), question levels (P< .0256), and student's ability (P < .0323). 'F' test also revealed a significant interaction between question types and question levels   (P< .0047), which indicated that remember - a - generality multiple-choice (RGM) group performed better (x = 2.6) than remember - a - generality essay RGE (x= 1.2), and remember - an - instance multiple-choice (RIM)   (x = 2.4) performed better than remember - an - instance essay (RIE)   (x =1.3), whereas the performance of use - a - generality groups didn't differ from the multiple-choice type questions to essay ones (See, Table 9).

Table(9). A Three-Way ANOVA summary table for question types, levels, and student's ability on a RIE sub-test .

Source	DF:	SS	MS	F-test	P value
Question type (A)	1	14.081	14.081	12.428	0.0006 **
Question level (B)	2	8.614	4.307	3.802	0.0256 *
AB	2	7.138	3.569	3.150	0.0470 *
Student's ability (C)	2	8.045	4.023	3.551	0.0323 *
AC	2	1.504	0.752	0.664	0.5171
BC	4	2.770	0.962	0.611	0.6555
ABC	4	2.395	0.599	0.528	0.7151
Error	102	115.56	1.133		

In terms of question <u>types</u>, Scheffe test showed at (.05) level of significance that groups who got multiple-choice type questions had higher mean than groups who got essay type questions (x = 2.06 vs.1.21). In terms of question <u>levels</u>, Scheffe test revealed that groups who got remember - a - generality level questions had a higher means (x= 2.6) than those who got remember - an - instance (x = 2.4) or use - a - generality (x = 1.1) levels questions. In terms of <u>ability</u> levels, the high ability students' mean was higher than the low ability students' one (x = 1.86 versus 1.09) (See, Table 10).

Table (10). Means, standard deviations, and number of students in each cell for question types, and levels on a RIE sub-test.

Q. types		Question levels			
		RI	RG	UG	TOTAL
Multiple	**x**	2.4	2.6	1.1	2
	(SD)	(1.5)	(1.1)	(.91)	(1.2)
	n	20	20	20	60
Essay	**x**	1.3	1.2	1.2	1.2
	(SD)	(1.1)	(.95)	(.81)	(.95)
	n	20	20	20	60
TOTAL	**x**	1.8	<u>1.9</u>	1.2	1.6
	(SD)	(1.3)	(1.1)	(.86)	(1.1)
	n	40	40	40	120

At the same time Scheffe test did not reveal any significant difference between medium and low student's ability (x=1.61 versus x=1.09) nor between medium and high ability students (x = 1.61 versus 1.86) (See, Table 11) .

Table (11). Means, standard deviations and number of students in each cell for question types, and student's ability on a RIE sub-test.

		Student's ability			
Question types		**high**	**medium**	**low**	**TOTAL**
Multiple	**x**	2.41	1.90	1.4	2.06
	n	29	21	10	60
Essay	**x**	1.31	1.3 20	0.81	1.21
	n	29		11	60
TOTAL	**x**	1.86	1.61	1.09	1.64
	(SD)	(1.2)	(1.3)	(.94)	(1.14)
	n	58	41	21	120

Results of the Remember - a - Generality Essay Sub-Test (RGE)

On the Remember - a - Generality Essay Sub-Test (RGE), general 'F' revealed a significant difference for student's ability only(P < .0001) (See, Table 12).

Table (12). A Three-Way ANOVA summary table for question types, levels, and student's ability on a RGE sub-test.

Source	DF:	SS	MS	F-test	P value
Question type (A)	1	4.06	4.06	1.78	0.184
Question level (B)	2	13.04	6.52	2.86	0.061
AB	2	5.78	2.89	1.27	0.285
Student's ability (C)	2	44.59	22.29	9.78	0.001 **

AC	2	2.48	1.24	0.54	0.581
BC	4	2.81	0.70	0.30	0.871
ABC	4	4.09	1.02	0.45	0.772
Error	102	232.41	2.27		

The Post Hoc ANOVA by using a Scheffe test at (.05) level of significance revealed that high ability students performed better than low ability ones (x =3.69 versus 2), whereas the medium ability students performed better than the low ability ones (x = 3.12 versus 2). At the same time, Scheffe test did not reveal any significant difference between the mean of high ability students and the mean of the medium one (x = 3.69 versus 3.12) (See, Table 13).

Table (13). Means, standard deviations and number of students in each cell for question types, and student's ability on a RGE sub-test.

		Student's ability			
Question type		**high**	**medium**	**low**	**TOTAL**
Multiple	**x**	3.5	2.71	2	2.97
	n	29	21	10	60
Essay	**x**	3.89	3.55	2	3.43
	n	29	20	11	60
TOTAL	**x**	3.69	3.12	2	3.20
	(SD)	(1.32)	(1.6)	(1.81)	(1.57)
	n	58	41	21	120

Results of the Use - a - Generality Essay Sub-Test (UGE)

On the Use - a - Generality Essay   Sub-Test (UGE), general 'F' revealed a significant difference for the question levels (RI, RG, and UG) only ($P < .0339$) (See, Table 14).

Table (14).   A Three-Way ANOVA summary table for question types, levels, and student's ability on a UGE sub-test.

Source	DF:	SS	MS	F-test	P value
Question type (A)	1	0.67	0.67	0.98	0.3235
Question level (B)	2	4.76	2.38	3.49	0.0339 *
AB	2	1.24	0.62	0.91	0.405
Student's ability (C)	2	3.32	1.66	2.44	0.0922
AC	2	0.63	0.41	0.61	0.5428
BC	4	2.62	0.65	0.96	0.4304
ABC	4	2.14	0.53	0.78	0.5359
Error	102	69.42	.68		

But the post hoc ANOVA by using Scheffe test at (.05)level of significance did not reveal any significant difference between the means of three question levels (RI, RG, and UG) ($x = 1.05$ vs. 1.45 vs. 1.35), though the mean of RG group was higher than the UG or RI ones (See, Table 15).

Table (15). Means, standard deviations, and number of students in each cell for question types, and levels on a UGE sub-test.

Q. types		Question levels			
		RI	RG	UG	TOTAL
Multiple	**x**	.95	1.35	1.35	1.21
	(SD)	.72	.82	.89	.81
	n	20	20	20	60
Essay	**x**	1.1	1.55	1.35	1.33
	(SD)	.69	.80	1	.83
	n	20	20	20	60
TOTAL	**x**	1	1.45	1.35	1.27

		.70	.81	.94	.82
	(SD)				
	n	40	40	40	120

Results of the Total Multiple-Choice Test

On the total multiple-choice test, general 'F' revealed a significant difference for the student's ability only (high, medium , and low) only ($P <$ .0001) (See, Table 16).

Table (16). A Three-Way ANOVA summary table for the total multiple-choice test.

Source	DF:	SS	MS	F-test	P value
Question type (A)	1	4.222	4.222	1.985	0.162
Question level (B)	2	9.611	4.806	2.259	0.109
AB	2	7.222	3.611	1.697	0.188
Student's ability (C)	2	65.20	32.60	15.32	0.0001 **
AC	2	3.574	1.767	0.840	0.4347
BC	4	8.005	2.001	0.941	0.4436
ABC	4	1.676	0.419	0.197	0.9394
Error	102	216.99	2.12		

The post hoc ANOVA by using Scheffe test at (.05) level of significance revealed significant difference between high student's ability versus the low ones (x = 7.759 versus x=5.714), and medium student's ability versus low ones (x = 7.024 versus x =5.714). But it failed to reveal a significant difference between the mean of high versus medium ability students (x = 7.759 versus x = 7.024) (See, Table 17).

Table (17). Means, standard deviations and number of students in each cell for the total multiple-choice test.

Question types		Student's ability high	medium	low	TOTAL
Multiple	x	7.75	6.57	5.5	6.96
	n	29	21	10	60
Essay	x	7.75	7.5	5.90	7.33
	n	29	20	11	60
TOTAL	x	7.75	7.02	5.71	7.15
	(SD)	(1.18)	(1.62)	(1.82)	(1.54)
	n	58	41	21	120

Results of the Total Essay Test

On the total essay test, general 'F' revealed a significant difference for the student's ability (high, medium, low) only (P< .0004) (See,   Table 18).

Table (18). A Three-Way ANOVA summary table for the total essay test.

Source	DF:	SS	MS	F-test	P value
Question type (A)	1	0.842	0.84	0.12	0.7216
Question level (B)	2	16.29	8.14	1.23	0.2951
AB	2	13.10	6.55	0.99	0.3739
Student's ability	2	113.2	56.6	8.58	0.0004 **
AC	2	1.593	0.79	0.12	0.8864
BC	4	9.883	2.47	0.37	0.8264
ABC	4	11.44	2.86	0.43	0.784
Error	102	672.9	6.59		

The post hoc ANOVA by using Scheffe test revealed significant difference between the levels of student's ability i.e., high versus low   (x = 7.017 versus x= 4.286) only. But it did not reveal any significant difference between high versus medium ability students (x = 7.017 vs. x = 5.817) nor between medium versus low (x = 5.817 versus x = 4.286) (See, Table 19).

 Table (19). Means standard deviations and number of students in each cell for the total essay test.

Question types		Student's ability			
		high	medium	low	TOTAL
Multiple	x	7.24	5.73	4.5	6.25
	n	29	21	10	60
Essay	x	6.79	5.9	4.09	6
	n	29	20	11	60
TOTAL	x	7.01	5.81	4.28	6.12
	(SD)	(2.2)	(2.67)	(2.94)	(2.6)
	n	58	41	21	120

The Results of the Overall Test

On the overall test, there were the following results:

1. General 'F' didn't show significant differences between essay versus multiple-choice questions (p >.760), thus, there was no need to use the Scheffe test. By this result, we answered the first question of this study "Which type of questions (essay versus objective) has more effect on students' overall learning?", by stating that there were no differences between essay and multiple-choice tests, though the mean of essay test was higher than the mean of multiple-choice (See, Table 20 &20:1 and Figure 1). At the same time, we accepted the first null hypothesis of this study which says "There are no significant differences at (0.05) a priori level of significance between the essay questions and the objective ones on all levels of learning (RI, RG, UG).

Table (20). A Three-Way ANOVA summary table for the overall learning test.

Source	DF:	SS	MS	F-test	P value
Question type (A)	1	1.292	1.292	0.093	0.7606

Question level (B)	2	48.21	24.17	1.740	0.1806
AB	2	20.41	10.20	0.737	0.4811
Student's ability (c)	2	349.6	174.8	12.62	0.0001 **
AC	2	9.441	4.721	0.341	0.712
BC	4	17.81	4.453	0.321	0.863
ABC	4	16.12	4.031	0.291	0.8832
Error	102	1412.7	13.85		

Table (20:1) shows the mean of each test type.

Test Types	**Essay**	**Multiple-choice**	**Total**
Mean	13.33	12.22	13.27

Figure (1) shows the mean scores of essay and multiple-choice questions on the overall learning test.

2. General 'F' didn't also show significant differences among the three levels of questions (RI, RG, and UG) (p = .180), thus, there was no need to use a Scheffe test. By this result, we answered the second question of this study "Which level of questions (RI, RG, and UG) has more effect on students' overall learning?", by stating that there were no differences among three of them, though the mean of RG group was higher than RI and UG groups (See,Table 22:2 and Figure 2). At the same time, we accepted the second null hypothesis which says "There are no significant differences at (0.05) a priori level of significance among the three levels of questions (RI, RG, UG) on all levels of learning.

Table (20:2) shows the mean scores of each level of questions.

Test levels	RI	RG	UG	Total
Mean	12.63	14.30	12.90	13.27

Figure (2) shows the mean differences between the three levels of questions (RI, RG, and UG) on the overall learning test.
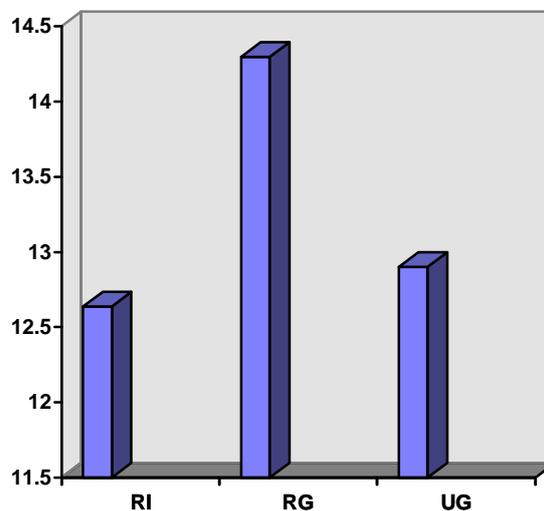
3. General 'F' showed significant differences among the three levels of student's ability (high, medium, and low) (p >.0001). The post hoc ANOVA (by using Scheffe test) revealed that there is a significant difference between high and low student's ability (x = 14.776 versus 10) and between medium and low student's ability (x = 12.841 versus 10). But it did not reveal any significant difference between high versus medium of student's ability means (x =14.776 vs. x = 12.841). By this result, we answered the third question of this study "Which level of student's ability (high, medium, or low) has more effect on the students' overall learning?, by stating that the performance of the high ability students was better than the performance of the medium and low

ability ones on the overall learning (See, Table 20:3 and Figure 3). At the same time , we rejected the third null hypothesis of this study which says "There are no significant differences at (0.05) a priori level of significance among the effects of the three levels of the student's ability (high, medium, and low) on all levels of learning.

Table (20:3) shows the mean scores of each level of student's ability.

Student's ability level	high	medium	low	Total
Mean	14.77	12.84	10.00	13.27

Figure (3) shows the performance of the three levels of student's ability (high, medium, and low) on the overall learning test.
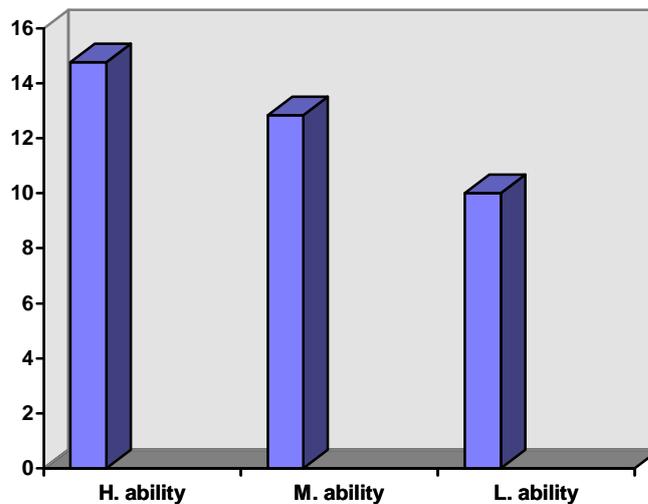
4. In terms of the interaction between types, levels and student's ability, general 'F' didn't show any significant interaction between question types (essay versus multiple-choice) and question levels (RI, RG, and UG) (p = .4811), nor between question levels (RI, RG, and UG) and student's ability (high, medium, and low) (p =.863), or between question types (essay versus multiple-choice) and student's ability (high, medium, and low) (p = .712). General 'F' didn't also show a significant interaction among question types

(essay versus multiple-choice), question levels (RI, RG, and UG) and student's ability (high, medium, and low) (p = .8832) simultaneously. By these results, we answered the fourth, fifth, sixth, and seventh questions of this study, they are **question 4**: "Is there an interaction between question types (essay versus objective) and question levels (RI, RG, UG) on students' overall learning ?", **question 5** "Is there an interaction between question levels (RI, RG, UG) and student's ability (high, medium, low) on students' overall learning ?", **question 6** "Is there an interaction between question types (essay versus objective) and student's ability on students' overall learning ?", **question 7**: Is there an interaction between types and levels of questions and student's ability on students' overall learning ?.

At the same time, we accepted the fourth, fifth, sixth, and seventh hypotheses, they are:

Hypothesis 4: There is no significant interaction at (0.05) a priori level of significance between question types and their levels on all levels of learning,

hypothesis 5 "There is no significant interaction at (0.05) a priori level of significance between question types and student's ability on all levels of learning", **hypothesis 6** "There is no significant interaction at (0.05) a priori level of significance between question levels and student's ability on all levels of learning", **hypothesis 7** "There is no significant interaction at (0.05) a priori level of significance between question types, levels, and students' ability on all levels of learning", by stating that there were no interactions among question types, levels and student's ability (See, Table 20:4).

Table (20:4) shows the interaction among types, levels of questions and student's ability.

Q. levels	RI			RG			UG			
S. ability	H	M	L	H	M	L	H	M	L	Total
Essay x	15	12.1	9.7	14.3	14.1	10.3	14.2	13.7	10	13.3

	n	10	6	4	9	8	3	10	6	4	60
Multiple	**x**	14.2	11.6	7.5	16.5	13.5	11.6	13.8	12	10.6	13.2
	n	9	8	3	11	6	3	9	7	4	60
Total	**x**	14.6	11.8	8.7	15.6	13.8	11	14	12.8	10.3	13.27
	n	19	14	7	20	14	6	19	13	8	120

Results of the Experimental and Control Groups

General F' showed significant differences between the experimental and control groups on the <u>total multiple-choice test</u> (P >.012). But Scheffe test didn't reveal any significant differences between each experimental group and the control group (See, Table 21).

Table (21). Means, standard deviations, and number of students in each cell of the experimental and control groups on the overall multiple-choice test.

Group	RIM	RGM	UGM	RIE	RGE	UGE	Control
x	6.45	7.5	6.95	7.45	7.7	6.85	6.15
(SD)	(1.95)	(1.1)	(1.66)	(1.43)	(1.59)	(1.75)	(1.34)
n	20	20	20	20	20	20	20

General 'F', on the other hand, didn't show any significant differences between the experimental and control groups on the <u>total essay test</u> (p > .2211) (See, Table 22), nor on <u>the overall test</u> (p > .1685).

Table (22). Means, standard deviations, and number of students in each cell of the experimental and control groups on the overall essay test.

Group	RIM	RGM	UGM	RIE	RGE	UGE	Control
x	5.72	7.45	5.6	5.65	5.95	6.4	<u>5.3</u>
(SD)	3.68	2.41	2.38	2.27	2.73	2.20	3.08
n	20	20	20	20	20	20	20

Thus, there was no need to use Scheffe test for post hoc ANOVA. By these results, we answered the last question of this study which says "Are there significant differences between the performance of the experimental groups which manipulated questions during the experiment compared with the performance of the control group which did not manipulate any questions during the experiment?", by stating that there were no differences among the means of all groups, though the mean scores of (RG) group, then the mean scores of (UG) was highest, whereas the mean scores of the control group was the lowest (See, Table 23 and Figure 4). At the same time, we accepted the eighth null hypothesis which says "There is no significant difference between the control groups' performance compared with the experimental groups' performance".

Table (23). Means, standard deviations, and number of students in each cell of the experimental and control groups on the overall learning.

Group	RIM	RGM	UGM	RIE	RGE	UGE	Control
x	12.17	14.95	12.55	13.1	13.65	13.25	11.5
(SD)	5.33	3.33	3.60	3.46	4.04	3.74	4.16
n	20	20	20	20	20	20	20

Figure (4) shows the six experimental groups and the control group on the overall learning test.
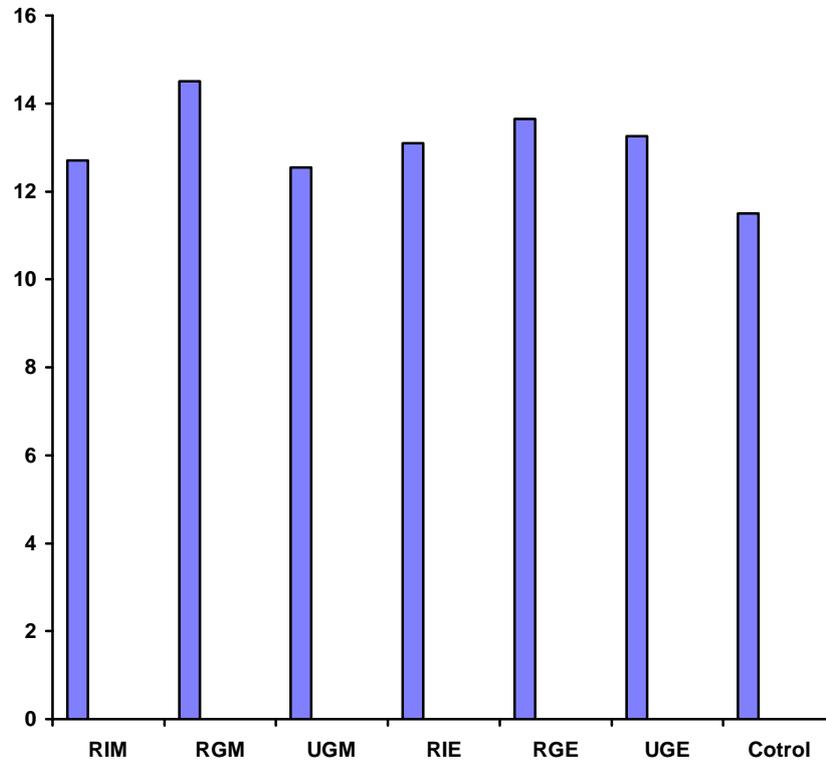
Chapter Five

Discussion, Conclusions, and Recommendations

Chapter Five

Discussion, Conclusions and Recommendations

The main aim of this study was to investigate the effect of question types (essay versus multiple-choice) and question levels: Remember - an - instance (RI), ýRemember a Generality (RG), and Use a Generality(UG) on three levels of learning (RI, RG, and UG) based on Merrill's Taxonomy . Thus, the discussion of the results of this study will take the same direction of its aim.

Before discussing the results of the study, it is worthwhile to mention that this study was limited to ʺMerrill's Taxonomyʺ: Remember - an - Instance (RI), Remember - a - Generality (RG), Use - a - Generality (UG). The Find - a - Generality (FG) level was not used because it has low relevance to most classroom objectives, and because ʺMerrillʺ has not yet adequately specified the nature of questions for the Find - a - Generality level (See, Darwazeh, 1982).

The results of this study were discussed according to the following items:

    1. Type of questions.

    2. Level of questions.

    3. Student's ability.

    4. Experimental versus control groups.

1. Type of questions:

The results of this study revealed somewhat that essay type questions have a greater effect in facilitating learning than multiple-choice ones. Experimental groups which received essay type questions during experiment performed better than those who received multiple-choice questions on most sub, post, and the total tests (See, Table 24).

Post-tests	Multiple-choice scores	Essay scores	P - Value
Remember Instance Multiple-choice	2.700	2.66	Non-sig.
Remember Generality Multiple-choice	2.233	2.43	Sig.
Use Generality Multiple-choice	2.033	2.23	Non- sig.
Remember - an - Instance Essay	2.067	1.21	Sig.
Remember - a-Generality Essay	2.975	3.43	Non- sig.
Use - a - Generality Essay	1.217	1.35	Non- sig.
Total Multiple-choice	6.967	7.33	Non-sig.
Total Essay	6.258	6.00	Non-sig.
Overall learning	13.225	13.33	Non-sig.

This trend of these results could be explained in that the essay type questions which asked students to write his/her answer in an open format and asked, in most cases, about general information require students to implement a deep mental process in understanding the text, hence to comprehend it more on different levels of learning. Whereas, the multiple-choice questions which asked students to check the right answer and asked, in most cases, about specific information <u>don</u>'t require students to implement a deep mental process in understanding the text, hence don't lead to comprehend it on different levels of learning. It required them to remember one level - a remembrance one instead.

This explanation were supported by Arrasmith, Sheehan & Applebaum 1984; Hammbleton & Merphy (1992); Lukhele, Thissen, & Wainer (1994); Roderick & Anderson (1968) when they found that essay questions have a <u>higher</u> effect on students' achievement than multiple-choice; and, on the other hand, oppose the results of Perkins (1984) and Duchastel & Nugester (1982) which found that multiple-choice questions have a greater effect on students' achievement than essay ones.

2. Level of questions:

The results of this study revealed, on most sub-tests and the total tests, that the remember a generality (RG) level questions first, then the use a generality (UG) level questions have a greater effect in facilitating learning than the remember - an - instance (RI) question level. In other words, the experimental groups which received (RG) questions or the (UG) questions during experiment performed better than those groups who received (RI) questions on most sub post -tests and on the total tests (See, Table 25).

Post-tests	RI	RG	UG	P - Value
RIM	2.629	2.950	2.475	Sig.
RGM	2.225	2.375	2.400	Non-sig.
UGM	2.100	2.275	2.025	Non-sig.
RIE	1.875	1.925	1.125	Sig.
RGE	2.763	3.325	3.525	Non-sig.
UGE	1.050	1.450	1.350	Non-sig.
Total Multiple-choice	6.950	7.600	6.900	Non-sig.
Total Essay	5.688	6.700	6.000	Non-sig.
Overall learning	12.637	14.300	12.900	Non-sig.

This trend of these results could be explained as that (RG) level questions which require students to remember the basic and general information, hence to comprehend the main ideas in the text would help students to retrieve information not just on the remember a generality (RG) level of learning but also on the use a generality (UG) level, because the (UG) level of learning could not be reached unless the student can master the (RG) level. The (UG) level of learning from "Merrill's" point of view means to use (apply) the learned general information in a new situation. Thus, the (RG) level is considered as a prerequisite of (UG) level.

This explanation agreed with the results of Al-Nayef (1989), Redfield (1981), and the two experiments which were conducted by Andre et al. (1980) when they found that high level questions such as RG and UG have greater effects on students' learning than low level questions such as RI level. On the other hand, these results disagreed with the results of Perkins et al. (1990) who found that there were significant differences for the factual questions (RI) but no significant differences between the generalization (RG)

and inference (UG) levels of questions. RI question level has a greater effect on students' learning than RG and UG question levels as they found.

These results could also be explained that giving students higher level questions such as RG and UG would help them to activate their cognitive strategies in a better way than giving them lower level question such as RI. This is why that the performance of RG, then the UG experimental groups performed better than the RI experimental group on most sub-tests and on the total tests (See, Table 25).

This result agreed also with what Rayan (1973); Redfield (1981); Richards and Vesta (1974); Royer and Konold (1984), found in their studies that UG questions have a greater effect on student' learning than RI ones. But, on the other hand, this result opposes to the findings of Felker and Dapra (1975); Perkins et al. (1990); Samson et al. (1987) which revealed that low - level questions (RI) had a greater effect on students' achievement than application questions (UG).

3. Student's ability

The results of this study revealed clearly and significantly that high student's ability then the medium student's ability performed better than the low student's ability on all sub-tests and the total test (See, Table 26).

Post-tests	H. ability	M. ability	L. ability	P-Value
RIM	2.825	2.683	2.286	Significant
RGM	2.552	2.268	1.857	Significant
UGM	2.379	2.073	1.571	Significant
RIE	1.862	1.610	1.095	Significant
RGE	3.698	3.122	2.000	Significant
UGE	1.457	1.085	1.190	Non sig.
Total Multiple	7.759	7.024	5.714	Significant
Total Essay	7.017	5.817	4.286	Significant
Overall learning	14.776	12.841	10.00	Significant

These results showed that high ability then the medium ability students performed better than low ability ones on all sub-tests and the total test. This is of course due to their high mental abilities which enable them to implement their cognitive strategies in understanding the text effectively and to comprehend the learned information easily, thus to be more effective in manipulating the text than the low ability students. This result supported by Darwazeh (1982); Memory (1983); and Sanders (1973) when they found in their studies that high ability students performed better than low ability ones on post tests measuring different levels of learning (RI, RG, UG).

In terms of interactions, results showed that there were no significant interactions between the three independent variables specially the two main variables: type and level of questions (See the results in chapter four).

4. Experimental versus control groups:

Although there were no significant differences between the performance of the experimental groups compared against the performance of the control

group (p=.16), but the mean scores showed that the performance of the control group, who did not receive any type of questions during the experiment was the lowest (x= 11.5) among all the groups who received different types of questions on different levels of learning (See, Table 27)

Group	RIM	RGM	UGM	RIE	RGE	UGE	Control
x	12.17	14.95	12.55	13.1	13.65	13.25	11.5
(SD)	5.33	3.33	3.60	3.46	4.04	3.74	4.16
n	20	20	20	20	20	20	20

In other words, the means of the experimental groups were higher than the mean of the control group though the differences were not significant. The direction of this result was supported by Darwazeh (1996) when she found that different levels of adjunct questions during instruction had activated and facilitated different levels of learning. This means that the use of adjunct questions during instruction have a greater effect than the use of the passage only without adjunct questions. According to this study, the researcher recommends teachers and book designers to present different adjunct questions during instruction, because this will help students understand the given text and be able to answer the post-test questions effectively.

Application of the study

= = = = = = = = = = = = =

The findings of the study will be of great importance for the following elements of the instructional process:

A. Students:

Students who are interested in questions and tests as a learning strategy will get benefit from this study as follows: Students will know that they shouldn't complain all the time of the type of questions that the teachers use in evaluating their achievement. The results of the study revealed that the two types of questions (essay vs. multiple-choice) have different effects on learning depending on the level of learning they want to induce. Their effects will vary from promoting remember levels to application levels of learning.

B. Teachers

Teachers can benefit from this study as follows:

1. This study should help teachers to weigh out their questions and students' abilities and instructional process aim before they decide on the type and level of questions they use for evaluating their students' achievement .

2. The results of this study will be used for educating teachers on how to use questions effectively either for teaching or for testing, and informing them that the RG questions are the basic questions if they want their students to learn not just on remember levels but also on higher levels such as application.

3. Informing teachers that the essay type questions are better than multiple-choice ones in promoting different levels of learning. Although the effects of the two types vary from one learning situation to another, from one learning level to another, and from student's ability level to another.

C. Exam designers and researchers

This study should help test designers and researchers in the following :

1. It shows test designers and researchers the importance of tests and exams in learning; exams are not only used to test students' achievement, but they can also use as a mean for providing students with new information.

2. Designers and examiners can benefit from this study in designing their tests. Particularly the General Secondary Examination (Tawjihi), by considering type, level, students' ability, and the instructional aims before putting these tests .

3. The researcher hopes that it would invite other interested researchers to build on the findings of this study to reach the ultimate goal of creating a reliable procedure for using questions effectively.

Recommendations
= = = = == = = = ==

1.Teachers are advised to use the RG level questions frequently in order to help students to understand the intended text and then to answer the higher level of questions later on.

2. Teachers are also advised to use the essay type questions frequently in order to promote students' cognitive strategies deeply, hence to increase their academic achievement on all levels of learning.

3. Researchers should

     a. conduct studies on each type and level of questions,

     b. apply further research directed to other topics matter rather than English,

     c . apply further research directed to other students rather than freshman college ones, and

     d. test different levels of questions by using "Bloom's Taxonomy" with its six levels, beside the Merrill's taxonomy.

4. The testing system should take into consideration the specific instructional objectives of the course to be measured in the light of the general educational taxonomies. Moreover, teachers should expose students to different types and different levels of questions with an emphasis on high cognitive level questions such as remember a generality and use a generality levels.

5. The Ministry of Higher Education should organize continuous training programs for teachers to increase their efficiency in evaluation and test construction.

6. The Ministry of Higher Education is strongly recommended to adopt a national testing center that would take care of all testing services including a testing bank to review the old tests and standardized them for benefit of students and interested researchers.

References of the Study

References
= = = = = = = =

Al-Nayef, M. A. (1989). <u>An Investigation of the Effect of Cognitive Questioning Level on the Reading Comprehension Achievement of Eleventh Grades</u>. Master thesis, Yarmouk University, Irbid.

Al-Makhzoomy, K.(1986). Teaching reading comprehension to secondary students in Jordan: Suggestions   improvement. <u>Dirasat</u> Vol. XIII, No. 6, pp. 19-29.

Al -Smadi H. A. (1992) . <u>Question types used by   teachers of English in evaluating the reading comprehension skill of the studies in the intermediate community colleges in Jordan and   the relationship between these types and   teachers' teaching experience and qualification</u> .Master thesis. Jordan : Jordan University .

Andre, T. (1979). Does answering higher level questions while reading facilitate productive learning ? <u>Review of Educational Research</u>,49, 280 - 313.

Andre, T. et al.(1980)."Adjunct application questions facilitate   later application or do they ?" <u>Journal of Educational   Psychology</u>, 72, 4, 533-543.

Anderson, R. C. & Biddle, W. B. (1975). On asking people questions about what they   are   reading. In G. Bower (Ed.), <u>Psychology of Learning and Motivation</u> . New York : Academic Press, vol. 9, pp. 89-132

Arrasmith, D. G., Sheehan, D. S. and Allebaum, W. R.(1984). A compari- son of the selected-response strategy and the constructed-response strategy for assessment of a third-grade writing task. Dallas, Department of Testing and Information Systems. <u>The Journal of Educational Research</u>, vol. 77, no. 3, pp.172-177.

Bloom, B. (ed.) (1956).Taxonomies of education objectives, <u>Handbook I: Cognitive Domain</u>, New York: Mckay.

Brock , C . A . (1986) . The effects of referential questions on ESL classroom discourse .TOSEL Quarterly, 20 (1) , 47-59 .

Brown, H. D. (1987). Principles of Language Learning and Teaching. 2nd edition. United States of America: Prentice-Hall, inc.

Bridgeman , B. (1989). Comparative validity of multiple-choice and free-response items on the Advanced Placement Examination in Biology (CBR No. 89-2). New York : College Entrance Examination Board.

Bridgeman , B.(1992). A comparison of quantitative questions in open-ended and multiple-choice formats. Educational Testing Service. Journal of Educational Measurement, vol.29, no. 3, pp. 253-271.

Bridgeman, B. and Rock, D. A. (1993). Relationships among multiple-choice and open-ended analytical questions. Educational Testing Service. Journal of Educational Measurement, vol.30, no. 4, pp. 313-329

Bridgeman, B. and Lewis, C. (1994). The Relationship of Essay and Mutiple-choice Scores with Grades in College Courses. Educational Testing Service. Journal of Educational Measurement. vol.31, no.1 pp. 37-50.

Darwazeh, A. N. and Reigeluth, C. M. (1982a). "Type and position of adjunct question: Their effects on memory and application", IDD & E Working Paper, no. 7. Syracuse- University, N. Y.

Darwazeh, A. N., (1982b). " Student - Generated Versus Teacher Generated Adjunct Questions: Their Effects on Remembrance and Application Level Learning", Doctoral dissertation , Syracuse University , N . Y .

Darwazeh, A. N.(1996). How can teachers activate cognitive strategies by using adjunct questions: A prescriptive instructional model for improving learning and teaching. Paper presented at the 43rd World Assembly for International Council on Education for Teaching. Amman, Jordan (Dec.16-21, 1996).

Darwazeh, A. N.(1996). Conditions affecting embedded versus generative cognitive strategy activators: Cross cultural comparisons. Bethlehem University Journal, vol. 15. Bethlehem, Palestine.

Duchastel, P. C. and Nungester, R. J. (1982). Testing effects measured with alternative test forms. USA: Pennsylvania. Journal of Educational Research, vol. 75, no. 5, pp.309-311.

Ebel, R. L. and Frisble, D. A. (1979). Essentials of Educational Measurement, (4th ed). Englewood Cliffs, N. J.: Prentice -Hall .

Felker, F. and Dapra, R. A. (1975). Effects of question type and position placements on problem solving ability from prose material. Journal of Educational Psychology, 67(3), 380-384 .

Frase, L. T. (1968). Retention of prose material. Journal of Educational Psychology, 59, 4, pp. 244-249.

Foos, P. W., and Fisher, R. P.(1988). Using tests as learning opportunities. Florida International University. Journal of Educational Psychology, vol. 80, no. 2, pp. 179-183.

Gronlund , N.E ,and Linn , R.L., (1990) . Measurement and Evaluation in Teaching, (6th ed.). USA, Macmillan Publishing Co., Inc.

Hader, N. A. (1991). Level of Classroom Oral Questioning of Arabic Language Teachers at the Basic Education Stages. Master thesis, Yarmouk University, Jourdan.

Hambleton, R. K. and Murphy, E.(1992). A psychometric perspective on authentic measurement. University of Massachusetts: Lawrence Erlbaum Associates. Applied Measurement in Education, 5(1),pp.1-16.

Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. Auburn University, Journal of Experimental Education, 62(2), pp. 143-157.

Hill, P.W. (1984). Evaluation in Education: an International Review Series. Oxford: Pergamon Press, vol. 8, no.3, pp. 181-200.

Hoeppel, F.(1981). " A taxonomy analysis of questions found in aiding skills developmental books used in Maryland Community College". Dissertation Abstracts International, vol.41.no.12, 5040-A.

Hopkins, K. D. and Stanely, J. C.(1981). Educational and Psychological Measurement and Evaluation, (6th ed). Englewood Cliffs, N. J.: Prentice Hall.

Hudgins, B., Dorman, J., & Harris, M. (1979). Some effects of adjunct questions on intermediate grade children with differing reading comprehension abilities. Journal of Educational Research, 72(5), 259-265).

Hunkins, F. (1969). "Effect of analysis and questions on various levels of achievement". Journal of Experimental Education, vol.38, no.1, pp.45-58.

Kniep, W. M . and Grossman, G. (1979). The effects of high level questions in competitive and cooperative environments on the achievement of selected social studies concepts . The Journal of Educational Research , 73 (2) , 82-85 .

Kubiszyn, T. and Borich, G.(1990). Educational Testing and Measurement (3rd ed.). London: Scott, Foresman and company.

Loring , R. M. (1986) .Questions used by teachers with skilled and less skilled readers. Dissertation Abstracts International, 47 (8), 2974-A.

Lukhele, R., Thissen, D. and Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. Educational Testing Service. Journal of Educational Measurement, vol. 31, no. 3, pp. 234-250.

Martin , et al . (1994) . Teaching Science for All Children . USA , Library of Congress Cataloging-in- Publication Data .

Martin, J. (1979). Effects of teachers higher-order questions on student process and product variables in a single-classroom study. Simon Fraser University. The Journal of Educational Research, pp.183-186.

Mehrens, W. A., and Lehmann, I. J. (1978). Measurement and Evaluation in Education and Psychology, (2nd ed). New York: Holt, Rinehart & Winston.

Memory, D. M. (1983). "Main idea pre-questions as adjunct aids with good and low average middle grade readers". Journal of Reading Behavior, 2, 15, p. 37-47.

Merrill, M. D. (1983). "The component display theory." In C.M. Reigluth (ed.). Instructional Design Theories and Models: An overview of their current status. N J: Lawrence Erlbaum Associates.

Perkins, K. (1984). An analysis of four item types used in testing EFL reading comprehension. USA: Southern Illinois University at Carbondale, RELC Journal, vol. 15 no. 2, DEC.

Perkins, -Kyle; et al. (1990). A level of question in ESL reading comprehension. (ED 322709, 1982-1991).

Perry, M., VanderStoep, S. & Yu, S. (1993). Asking questions in first-grade mathematics classes: Potential influences on mathematical thought. American Psychological Association. Inc. Journal of Educational Psychology, vol. 35, no. 1, pp. 31-40.

Rayan, F. L. (1973). "Differentiated effects of levels of questioning on student achievement." Journal of Experimental Education, 41 (Spring 1973), pp. 63-67.

Redfield, D. L.(1981). "A meta-analysis of experimental research on teacher questioning behavior". Journal of Educational Research, vol. 51, no. 2, pp.237-245.

Redfield, D. L. & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. Review of Educational Research, 51, (2), 1105-A.

Rickards, J. P. ,& Di Vesta, F. J. (1974). Type and frequency of questions in processing textual material. Journal of Educational Psychology, 66, pp. 354-362.

Rinser, G. P. (1987)." Cognitive levels of questioning demonstrated by test items that accompany selected fifth-grade science textbooks" (Eric Document no.291752,1987).

Rodreck, M. C. & Anderson, R. C. (1968). "A programmed introduction to psychology versus a textbook-style summary of the same lesson". Journal of Educational Psychology, 54, pp. 381-387.

Royer, J. and Konold, C.(1984). Learning from text: methods of affecting reader intent. In Reading in a Foreign Language . James Alderson &

1st. Urquhart (Eds.) , London, Longman, pp.65-82.

Sanders, J. R. (1973). Retention effects of adjunct questions in written and aural discourse. Journal of Educational Psychology, 56(2), pp. 181- 186.

Samson, G. E. et al. (1987). The effects of teacher questioning levels on student achievement: A quantitative synthesis. The Journal of Education Research , 80 (5) , 290-295 .

Shavelson, R. J. et al. (1974). Effects of position and type of question on learning from prose material: Interaction of treatments with individual difference. Journal of Educational Psychology, 66(3), pp.40-48.

Shohamy, E.(1985). A Practical Handbook in Testing for the Second Language Teacher. Israel: Tel-Aviv University, experimental edition.

Thissen, D., Wainer, H., Wang, X-B.(1994). Are tests comprising both multiple-choice and free-response items necessarily less unidimensional than multiple-choice tests? An analysis of two tests. Journal of Educational Measurement, 31, pp. 113-123.

Tollefson, W. (1989). A system for improving teachers questions. USA: University of Washington Seattle. English Teaching Forum, vol.27, no. 1, pp. 6-9.

University English I(1993). "The Common Cold". Department of English, An-Najah National University, Nablus.

Watts, G. H., & Anderson, R.C. (1971). Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 62 (5), pp. 387-394.

Williams, J. P. (1963). "Comparison of several response   modes in a review program". Journal of Educational Psychology, 54, pp. 253-260.

Winne, P.H. (1979). Experiments relating teacher's use of higher cognitive questions to student achievement. Review of Educational Research, 49 (1), pp. 13-50.

Zeidner, M. (1990). "College Students' Reactions Towards Key. Facets

Classroom Testing". Assessment and Evaluation in Higher Education, 15, pp. 151-169 .

المراجع العربية

دروزة ، افنان نظير(١٩٨٣ـ١٩٨٤) أثر مستوى ومكان الأسئلة التعليمية من النص المدروس على ثلاثة مستويات في التعلم :التذكر والتطبيق والاكتشاف، مركز التوثيق والأبحاث، جامعة النجاح الوطنية.

دروزة ، افنان نظير(١٩٨٧) . الاسئلة التعليمية والتقييم المدرسي،ط١ . مكتبة خالد بن الوليد ، نابلس ، فلسطين .

دروزه ، افنان نظير(١٩٩٢) . دراسات تجريبية حول منشطات استراتيجيات الادراك كما يشتقها المدرس وكما يشتقها الطالب كأساس نظري لنموذجين تعليميين خاضعين للتجريب.مجلة جامعة بيت لحم ، ع ١١+١٢ ، ص ص: ٢٩ـ ٨٥ ، فلسطين.

دروزه ، أفنان نظير (١٩٩٥). إجراءات في تصميم المناهج ، ط٢، مركز التوثيق والمخطوطات والنشر، جامعة النجاح الوطنية، نابلس.

دروزه ، افنان نظير(١٩٩٧) .النظرية في وضع الأسئلة التعليمية "دليل المعلم والطالب" ، ط١ ، مكتبة النصر ، نابلس ، فلسطين.

دروزه ، افنان نظير(١٩٩٧) . الاسئلة التعليمية والتقييم المدرسي ، ط٢ ،مكتبة الفارابي، نابلس ، فلسطين.

دروزه ، افنان نظير(١٩٩٢). النظرية والبحث في التدريس. رابطة الجامعيين: دائرة البحث والتطوير، الخليل، الضفة الغربية.

Appendix B

The post-passage questions which were used in the experiment of this study

Number of the booklet (   )

<u>General Instructions</u>

1- Don't write your name on the booklet.

2- It is important that you remember your booklet number it appears on the top right side of this sheet.

3- You have <u>25 </u>minutes <u>only</u> to read the passage once.

4- You have <u>10</u> minutes <u>only</u> to answer the post questions.

5- Hand your booklet to your proctor when you finish.

6- The total time allowed for the passage and the questions is limited to <u>35</u> minutes.

Thanks

Remember - an - instance multiple-choice post passage questions

Circle the most appropriate answer for each of the following questions.

1- The winner of the 1954 Nobel prize for chemistry was
    a. Dr. Alian Bombard .   b. Marie Curie .
    c. Dr. Linus Pauling .   d. Qasim Amin .

2- Scientists have found .... kinds of viruses that cause cold .
    a. 150   b. 100
    c. over 200   d. less than 200

3- Common cold can be prevented by taking
    a. vitamin B   b. a protein
    c. a vaccine   d. lemonade

4- Scientists discovered in the .... that viruses cause cold.
    a. 1950s   b. 1970s
    c. 1940s   d. 1930s

5- If you take .... grams of vitamin C , it will prevent colds.
    a. 1 to 3   b. 2 to 3
    c. 1 to 2   d. 2 to 6

6-Which disease of the following has the same symptoms to the cold ?
    a. Influenza   b. Fever
    c. Cancer   d. Heart disease

7- Colds are caused by a
    a. germ   b. bacterium
    c. fungus   d. virus

<u>Remember -an - instance essay post passage questions</u>

Answer the following questions by writing your answer in the specified spaces .

1- Who was the winner of the 1954 Nobel prize for chemistry?

_____

_____

2- How many kinds of viruses that cause colds have scientists found ?

_____

_____

3- Common cold can be prevented by injecting the patient with_____

_____.

4- When did scientists discover that viruses cause colds ?

_____

_____

5- How many grams of vitamin "c" will prevent cold ?

_____

_____

6- What disease has similar symptoms to the common cold ?

_____

_____.

7- What causes cold , a virus or a bacterium ?

_____

_____.

Remember - a - generality multiple-choice post passage questions

Circle the most appropriate answer for each of the following questions:

1- The major symptoms of colds are
      a. headache, sleepiness and coughing.
      b. sneezing, headache and coughing.
      c. sneezing, yawning and headache.
      d. nausea, coughing and fever.

2- A vaccine is a
      a. liquid which is injected into the arm.
      b. tablet taken by mouth to cure a disease.
      c. disease which spreads by touch.
      d. disease developed by injection.

3- The major difference between cold and influenza is
      a. sleepiness.   b. tiredness.
      c. fever.   d. eating.

4- Two ways to avoid caching colds are avoid
      a. getting chilled and being wet in summer.
      b. touching people and shaking hands with others.
      c. eating ice-cream and drinking lemonade.
      d. watching television and smoking cigarettes.

5- Most cold viruses are spread through
      a. eating and drinking.
      b. hugging and kissing.
      c. hands and touch.
      d. coughing and sleeping.

6- A contagious disease is a disease which
      One. doesn't have treatment.
      Two. spreads by touch.
      Three. affects only pregnant women.
      Four. causes death.

Remember - a - generality essay post passage questions

Answer the following questions by writing your answer in the specified spaces:

1- Name three symptoms of the common cold.

    1- _____.

    2- _____.

    3- _____.

2- What is a vaccine ? Define it in your own words within one line.

_____.

3- What is the major difference between cold and influenza ?

_____.

4- What should you do to avoid catching cold ? Mention two ways.

    1_____.

    2- _____.

5- How do most cold viruses spread ? Mention two ways.

    1-_____.

    2-_____.

6- What is a contagious disease ? Define it in your own words within one line.

_____.

Use - a - generality multiple-choice post passage questions

Circle the most appropriate answer for each of the following questions:

1- You cannot take a vaccine for one time to prevent cold, but you can take it to prevent
  a. headache .   b. stomach ache .
  c. cholera .   d. blood pressure .

2- Medicine can be giver to patients in different compositions like injection. Another common way of taking medicine is
  a. tablets.   b. smelling air.
  c. drinking water.   d. tasting sugar.

3- People believe that you can shake hands with some one who has a headache because headache
  a. affects only old people. b. affect only young people.
  c. is a contagious disease. d. is not contagious.

4- Doctors say that influenza is a contagious disease as well as
  a. headache.   b. malaria.
  c. stomach ache.   d. blood pressure.

5- Skin infection is a
  a. disease.   b. medicine.
  c. symptom of a disease. d. fever.

6- Medicines cannot cure Sameer from influenza forever because influenza is caused by
  a. many kinds of viruses. b. one kind of viruses.
  c. many of bacteria.   d. one kind of bacteria.

Use - a - generality essay post passage questions

Answer the following questions by writing your answer in the specified spaces.

1-You cannot take a vaccine for one time to prevent cold. Can you take it to prevent cholera ? Yes ..... . No .....
Why ? _____
_____.

2- Medicines can be given to patients in different compositions like injections. Name another way of taking medicines that is commonly used by people.

_____
_____.

3- Do you think that you should or should not shake hands with some one who has a headache ? Yes ..... . No ..... .
Why ? _____
_____.

4- Influenza is a contagious disease. Do you think that malaria is a contagious one ? Yes ..... . No ..... .
Why ? _____
_____.

5- Do you think that skin infection is a disease or a symptom of a disease ? Explain you answer within two lines.

_____
_____.

6- Sameer has influenza and he is taking a medicine. Do you think that this medicine can cure him forever ? Yes .... . No
Why ? _____
_____.

Appendix C

**The achievement post test which was used
in the experiment of this study**

Tawjihi average (   )   **Booklet no. (   )**

<u>General Instructions</u>

1- Don't write your name.

2- Write your booklet number on the upper right side.

3- Write your Tawjihi average on the upper left side.

4- Answer all test questions within <u>30</u> minutes <u>only</u>.

5- Hand your test to the proctor when you finish.

Thanks

I. **Answer the following questions by writing your answer in the specified spaces:**

1- Who was the winner of the 1954 Nobel prize for chemistry?

_____.

2- How many kinds of viruses that cause cold have scientists found ? _____.

3- Common cold can be prevented by injecting the patient with a _____.

4- When did scientists discover that viruses cause colds ?

_____

II. **Circle the most appropriate answer for each of the following questions.**

1- If you take ...... grams of vitamin C , it well prevent colds.
    a. 1 to 2   b. 2 to 3
    c. 1 to 3   d. 2 to 6

2- Which disease of the following has the same symptoms to the common cold ?
    a. Influenza   b. Fever
    c. Cancer   d. Heart disease

3- Colds are caused by a
    a. germ .   b. bacterium .
    c. fungus .   d. virus .

III. Answer the following questions by writing your answers in the specified spaces :

1- What should you do to avoid catching cold ? Mention two ways.

    a. _____.

    b. _____.

2- How do most cold viruses spread ? Mention two ways.

    a. _____.

    b. _____.

3- What is a contagious disease ? Define it in your own words within one line.

_____.

IV. Circle the most appropriate answer for each of the following questions:

1- The major symptoms of colds are

    a- headache, sleepiness and coughing.

    b- sneezing, headache and coughing.

    c- sneezing, yawning and coughing.

    d- nausea, coughing and fever.

2- A vaccine is a

    a- liquid which is injected into the arm.

    b- tablet taken by mouth to cure a disease.

    c- disease which spreads by touch.

    d- disease developed by injection.

3- The major difference between cold and influenza is in the amount of

    a- sleepiness .   b- tiredness .

    c- eating .   d- fever .

V. Answer the following questions by writing your answers in the specified spaces:

1- Do you think that you should or should not shake hands with some one who has a cough ?Yes   No
   Why?
   _____.

2- Influenza is a contagious disease .Do you think that AIDS is a contagious disease one ?Yes ... No ...
   Why ?
   _____.

3- Do you think that blood pressure is a disease or a symptom of a disease ?Yes ...No...
   Why?
   _____.

VI. Circle the most appropriate answer for each of the following questions:

1- You can take a vaccine to prevent measles because
   1. it is a contagious disease.
   2. it is caused by one virus.
   3. it is caused by many viruses.
   4. it spreads by smelling.

2- Medicines cannot cure Ahmad from flu forever because flu is caused by
   a. many kinds of viruses.   b. one kind of virus.
   c. many kinds of bacteria.   d. one kind of bacteria.

3- Medicines can be given to patients in different compositions like injections. Another common way of taking medicines is
   a. eating.   b. sleeping.
   c. capsules.   d. smelling.

Appendix A

The passage which was used in the experiment of this study