AN INVESTIGATION INTO FIXED RESPONSE QUESTIONS IN SCIENCE AT SECONDARY AND TERTIARY LEVELS

By

ABDULLAH KHAMIS ALI AMBU - SAIDI

BSc. (Science Education) Sultan Qaboos University – Oman

MSc. (Science Education) Warwick University – U.K.

A thesis submitted in part fulfilment of the requirements for the degree of Doctor of Philosophy (Ph.D.) in Science Education

UNIVERSITY OF GLASGOW

FACULTY OF SCIENCE

CENTRE FOR SCIENCE EDUCATION

IN THE NAME Of ALLAH, THE MERCIFUL, THE COMPASSIONATE

To my wife *Muna*, and my son *Qusai*

ABSTRACT

This thesis focuses upon the effect of some factors on pupils'/students' responses in some assessment methods mainly fixed response questions.

The major aim of the research is trying to find out these factors in order to improve the assessment process.

The research was carried out over three years and the samples were selected from secondary schools both in the Glasgow area and in the Central Region of Scotland and the first year University of Glasgow students who were taking biology and chemistry courses.

The results of the first investigation which was carried out to find out the effect of changing the positions of the key answer and the most plausible distractor (strong distractor) in the multiple choice questions (MCQs) showed that it can have a significant effect in the facility value of the question. By placing the most plausible distractor next to the key answer, students can make an effective discrimination which leads more often to the correct answer. One factor, which may be the cause of the difference in the facility value of the MCQs, has an "optical" origin. This factor was examined at the school level. The results of this examination revealed that the differences in the facility values of the MCQs disappears or is at least reduced when suitably treated.

Because of the above differences in the facility values of MCQs, another fixed response method is offered which is called Structural Communication Grids (SCGs). However, this method was also investigated like the MCQs to find out if changing the positions of the options within the grid affect the facility value of the question. The results showed that this also brings, to some extent, a significant difference in the facility value of the grid questions but not to the some extent as the multiple choice questions.

In the same area (i.e. related to the grid questions), pupils' ranking in two methods of scoring the grid questions was examined. The results showed that there was no significant difference between the two methods.

The research is not only looking at the effects of some editing factors on the facility value of some fixed response questions but also to investigate the effects of some

psychological factors. The results showed that field - independent pupils perform better than field - dependent pupils in the fixed response questions based on the grid and open response questions. On the other hand, it has been found that there is no significant difference in performance between convergent and divergent pupils in multiple choice questions (MCQs) and structural communication grids (SCGs).

Another investigation was carried out to find out the relationship between pupils' knowledge structures and multiple choice test structure. The results revealed that the option in the question, which is closer to the concept on which the stem of the question is based, is more likely to be selected than other options.

The final investigation was conducted to find out how pupils' performances in the word association test (WAT) is related to their achievement in the multiple choice questions. The results showed that there is a positive correlation between pupils' performances in WAT and their achievement in the multiple choice questions (MCQs). The facility value of the MCQs of those pupils who are more able to associate concepts in WAT is higher than the facility value of the MCQs for those pupils who are less able to associate concepts in WAT. Furthermore, in most cases it was found that the options, which have high response frequencies in WAT, as more likely to be selected among other options in multiple choice questions.

ACKNOWLEDGEMENTS

Doing research in Science Education is not an easy task. It requires a lot of work, co-operation and help from different people and institutions. Consequently, this study does not only belong to me but also to those without whose generous help and support this research could not have succeeded.

At the beginning, I would like to express my appreciation and thanks to my supervisor, Professor Alex H Johnstone at the Centre for Science Education - University of Glasgow for his help, encouragement and advice which I received during my research. His suggestions and comments were indispensable for the completion of my study.

My special thanks go to Dr. Liz Leonard at the Teaching and Learning Service – University of Glasgow for her guidance, help and enthusiasm during my study. Her help was essential, that allowing me to apply my test materials at the tertiary level.

I also wish to acknowledge Dr. Fakhir Al – Naeme who gave constant help and guidance not only about this study but for everything. My thanks also to Dr. Norman Reid for his friendship, suggestions and help during my research. I am also thankful to my colleagues at the Centre for Science Education especially, Mr Ali Al – Shuaili and Mr Ghassan Sirhan for their invaluable friendship and support.

I would like to express my sincere gratitude to all the schools in which I worked and to all the teachers and pupils at Stirling High School, Larbert High School, Wallace High School, Balfron High School, St. Columba's High School and Cleveden High School for participating and helping with the work on which my research relies, and without which it would have been totally impossible to complete. Also my thanks go to the students and the lecturers of the Institute of Biomedical and Life Sciences and the Department of Chemistry – University of Glasgow for participating in my research at the tertiary level.

My thanks also to the Sultan Qaboos University, Sultanate of Oman for the financial support I received which allowed me to complete my study at Glasgow.

Finally, I am greatly indebted to my wife Muna and my son Qusai in Glasgow and my father, brothers and sisters back in Oman for the love, faith and patience they have invested in my efforts and without whom I would surely never have succeeded in performing my research.

CONTENTS

CHAPTER ONE: INTRODUCTION

CHAPTER ONE: INTRODUCTION

1.1 Introduction:

Assessment is a process for obtaining information that is used for making decisions about students, curricula and programmes, and educational policy (Capper, 1996). It is a way that helps people to discover the truth about any educational system (Rowntree, 1987). It is like a window into the student's mind that gives opportunity for teachers to see what students know (Lorsbach *et al.*, 1992). Moreover, it is a central component and integral part of the curriculum (Kellington *et al.*, 1980; Wills, 1993; McGrath *et al.*, 1992).

A good teaching method needs a constant flow of information about pupils' progress or about their failure (West, 1993; Wisker, 1997; Mavrommatis, 1997). Therefore, assessment has been used for many years as the tool for reflecting the success and failure of any school programme and gives feedback that helps to conduct effective teaching.

Many people think that assessment is something done at the end of the course or unit; however, it is part of the teaching and learning process and must be planned very well (Lloyd - Jones, 1986; Brooks, 1993). There should be a direct connection between the type of assessment used and what has been taught and learned in the school programme. This linkage is necessary for building an effective environment for learning in any institution (Burton *et al.*, 1997).

One may ask why schools or universities put a lot of effort in carrying out an assessment. Actually, different sectors in society rely on the results of the assessment process. Therefore, assessment should really reflect what is required or expected from it. There are many purposes that an assessor is seeking from his/her assessment. Some of these are important to students, others to institutions including teachers and some to society including the parents.

a. Purposes Concerning Students:

Assessment helps students to:

- obtain a qualification for entry into the next stage or level of education or the world of work (Skevington, 1994).

- support and guide them in their own learning (Black, 1993; Balla *et al.*, 1994; Baumann *et al.*, 1997).
- motivate them and give incentives for learning (Knutton, 1994).
- give them feedback about the progress they have made in their learning. As a result of this, individuals are able to identify their strengths, weaknesses and any difficulties they faced in their learning (Stewart - Tull, 1970; Hodson, 1986; Bailey, 1999).

b. Purposes Concerning Institutions (Including Teachers):

Assessment helps institutions and teachers to:

- allocate students to different sets or groups.
- give feedback to the teacher about his teaching effectiveness and meeting the objectives he desires to achieve (Abdullah *et al.*, 1997; Baumann *et al.*, 1997).
- decide what the next step in a curriculum plan or the type of work that teacher should do for a particular group or people (Gott *et al.*, 1995).
- measure and control standards.
- keep a record for every single student that may be needed by anyone.
- reveal errors and misconceptions.
- give students a grading or certificate.

c. Purposes Concerning Society (Including Parents):

Assessment helps society to:

- select the appropriate person for the appropriate work or career opportunity and provide a license to practice (e.g. Vets, Dentist, Teachers).
- report to the parent about their children progress.
- help in the choice of courses for the next stage in education.

Assessment can be conducted by different methods to gather information about students' performance. Unfortunately, many of these methods which have been used over many years are largely a measure of the ability to recall facts or to measure the lowest levels of understanding. Furthermore, some of these methods, if they have been used in an inappropriate way or if they do not measure desired outcomes could harm students (i.e. may give student less marks although he knows the content). Consequently, these methods should be selected, with much care, according to what is intended to be assessed. Unless the assessor has clear and understandable criteria about his/her assessment, it is unlikely that the teacher makes the right decisions about when, and what, and how, and by whom the assessment will be (Shaw *et al.*, 1993).

It has been found that a lot of time and money is spent on the process of assessment but, in some cases, without any benefit that leads to an improvement in the teaching and learning process. In other words, although assessment is a complex and powerful process in the classroom, some teachers are unaware about this and they do not use it effectively. Very little is obtained from the assessment methods that helps teachers to teach. Instead, students are given marks or grades, saying how 'good' or 'bad' they are, and putting them in an order of merit. Teachers should be aware of the potential of the assessment for the improvement of teaching and learning and its side effects if it is used in a wrong way.

Recently, there have been internal and external pressures for institutions to review and improve their practice in the area of assessment (Balla *et al.*, 1994). For example, there is considerable concern about assessment in England and Wales since the introduction of the national curriculum (Swain, 1989; Swain, 1991). There are several reasons for this pressure and they can be summarised as follows:

- Much assessment of the students is carried out, consciously or unconsciously, in so many different ways and with so many different purposes (Jackson, 1988). This assessment has a lot of weight and affects the future of the students. Therefore, effective assessment methods are needed.
- The big movement in the learning and teaching process is resulting in new methods of teaching. These methods, such as problem solving, help to promote effective learning and teaching. The assessment process should parallel any development and improvement in teaching and learning.
- The changing aims of higher education such as mix of academic, vocational and general transferable skills.
- Emphasizing deep learning rather than surface (shallow).
- To save staff time.

However, when we look to the process of assessment it is still carried out by the same methods as it has been for many years. It is, therefore, important that rapid changes in assessment should happen to keep up with what is happening in teaching and learning.

Lloyd - Jones (1986) reports that education, and of course assessment which is part of it must respond to the changing needs of society, to new branches of knowledge, to changing perceptions of its role and to refinement in its techniques.

1.2 The Problems of Assessment:

Carrying out any assessment process is not an easy task. It faces many problems and difficulties. Teachers should be familiar with these problems in order to eliminate or at least try to reduce them. Some of these problems are listed by Entwistle *et al.* (1992):

- Assessment systems are often difficult for departments to change as they depend on institutional or faculty regulations.
- It is difficult to establish a workable assessment system that balances ease of use with the promotion of desired learning activities and outcomes.
- In large first year classes (university level) the workload implications for some desirable assessment procedures make them unworkable in practice.
- The need to award classified degrees makes collaborative work difficult to assess fairly.
- Attempts to include innovative methods of assessment, or different skills, make it difficult to ensure that standards are maintained.

Furthermore, there are other problems mainly related to the assessment methods such as:

- Choosing a valid assessment method. In other words, ensuring the test is measuring what it is supposed to measure.
- Marker reliability, especially if more than one marker is involved.
- Using valid statistical methods and drawing valid inferences from the measure.

Nevertheless, these problems should not be obstacles to any improvement. A change in the current practices of assessment is necessary so that the assessment methods are no longer focused on the lowest levels of learning and they can provide teachers with insight into their methods and their tools of assessing and teaching.

In the light of the problems associated with assessment, this project seeks to examine in some details the strengths and weaknesses of some forms of objective testing.

CHAPTER TWO: LITERATURE REVIEW (1)

CHAPTER TWO: LITERATURE REVIEW (1)

2.1 Introduction:

Until now there is no one perfect or ideal method of assessment. Each method has its advantages and disadvantages. However, the suitable method of assessment to be used depends mainly on the outcomes required to be assessed and therefore teachers have to think deeply before choosing any assessment method (Lloyd - Jones, 1986). Balla *et al.* (1994) argue that good practice in assessment of students' performance is associated with selection of the method which matches the purpose of the assessment, the properties or characteristics being assessed and the objectives (intended outcomes) of instruction.

There are different classifications of classroom assessment made by educators (Gronlund, 1971; Marshall *et al.*, 1971; Thyne, 1974). Each classification is looking from a different point of view. Some of these classifications are as follows:

Classification 1: Selection Questions (True/False (T/F), Multiple Choice (MCQs), Structural Communication Grids (SCGs), Matching, Venn Diagrams) versus Supply Questions (Essay - short answer, Essay - extended answer, Completion).

Classification 2: Objective Questions (*Selection Type* such as True/False (T/F), Multiple Choice (MCQs), Structural Communication Grids (SCGs), Matching, Venn Diagrams and *Supply Type* such as Completion and Short Answer) versus Essay Questions (Restricted and Extended).

Classification 3: Fixed Response (Select from given answers) versus Open Response (Examinee provides answers).

The **key question** which could be asked here is that "*What are the problems of the current fixed response and open response methods of assessment?*". The following sections were written to answer the above question. The aim of this chapter is to give a background review of some of fixed response and open response methods by giving the descriptions, advantages and disadvantages of each method.

2.2 Assessment Methods:

2.2.1 Fixed Response Questions:

It is difficult to see the results of enhancing new teaching methods in students learning using conventional assessment methods such as some forms of objective testing

(Gibbs, 1995). However, there has grown up in recent years the use of many types of fixed response (objective) tests and essay type formats in a great many subject areas (Jeans *et al.*, 1978; Zeidner, 1987).

Objective testing can be defined as a test which is free from any subjective bias - either from the tester or the marker (Currie, 1986a; Leonard *et al.*, 1996). However, we have to be very clear about the word objective. In objective testing, the objectives that are intended to be tested are subjectively chosen (chosen by the teachers). The questions themselves have been subjectively written to match these objectives (testing the objectives from the teacher's point of view - professional judgement). The wordings are subjective (chosen by the teacher to test the objectives), and the interpretation of the scores is subjective (the results are subjectively judged by the examiner, to be satisfactory or otherwise, and users of the results, such as parents and employers). The only objective part of objective testing is the scoring because machine scoring can be used.

Nonetheless, fixed response (Selection Type of objective testing) can take various forms such multiple choice questions, true - false questions, matching questions, Venn diagrams and structural communication grids. The following paragraphs are going to discuss these types of objective testing.

2.2.1.1 Multiple Choice Question (MCQs):

i. Introduction:

Multiple choice questions (MCQs) are one of the main types of fixed response questions. They are generally recognized as the most widely applicable and useful type of objective testing (Gronlund, 1971). They have been used widely as the main assessment method to assess students' performance. They reached the peak of their popularity in the 1950s and 1960s because of their high reliability (White *et al.*, 1992). The main aim of introducing such questions was to replace the essay questions in many science examinations and to meet the social needs of an expanding society.

MCQs were derived from traditional intelligence testing (Black, 1998). Then, a further justification has been made for this type of question by the prevailing emphasis on managing learning through specification of behavioural objectives.

However, in the late 1960s criticism of these questions had begun in the United States where for the first time these questions were introduced (Black, 1998). The main criticism was that this type of question led pupils and teachers to think of learning as the art of picking the right answer. They were unable to test activities that would be more constructive, creative and holistic. Furthermore, the critics argued that these questions could not be the instrument to improve education because:

- they could only sample a narrow range of types of performance.
- they were not linked to schedules of the curriculum.

Other criticisms of the multiple choice questions came from the learning point of view. Learning is a complex process, which cannot be reduced to a routine of selection of a small component by using multiple choice type questions.

Despite the criticisms made above to the MCQs, they still play an important role in assessing students' academic attainment in the science curriculum. Many schools and universities use these types of questions to assess their students' performance all over the world and in different disciplines.

ii. Descriptions of Multiple Choice Questions:

MCQs are a series of questions (called items) each having a pre-determined correct answer (Bannister *et al.*, 1967; University of London, 1976). Each item (Gronlund, 1971; Johnstone *et al.*, 1972) consists of:

a. The stem: where the problem may be stated in the form of a direct question or an incomplete statement.

b. The options: where the solution is found. The options can be words, numbers, symbols, or phrases. The right option is called the key answer and the wrong options are called the distractors. The word distractor means that the wrong option distracts those students who are in doubt about the correct answer.

Example of MCQs from Biology:

In which animal does blood flow through vessels from the respiratory organ back to the heart before circulating through the rest of the body? ← **The Stem**

a. frog.

b. shark.

c. leeches. ← **The options**

d. snails.

e. termite.

Note: Option **a** is the key answer whereas options **b**, **c**, **d**, **e** are the distractors.

iii. Advantages of the Multiple Choice Questions:

Educators still do not reach a similar view about MCQs. Some of them strongly support the use of MCQs to assess students' performance while others are strongly against using them in any assessment process especially in higher education. Both groups have their own arguments. However, some teachers and educators prefer using MCQs to assess their students for several reasons.

a. A wide selection of topics can be tested at the same time and in a short period (University of London, 1976; Tamir, 1990; Dixon, 1994; Ndalichko *et al.*, 1997). The teacher can test widely the curriculum syllabus using these questions. This is because these type of questions only require from the examinee to select his/her answer from the given options. A study by Blum *et al.* (1986) found that students preferred to answer MCQs than essay type questions due to the time pressure. Students like to answer MCQs because they can answer more questions than in essay type questions in the allocated time of the exam.

b. They have the potential for testing large numbers of students in a short time (Daniels, 1968; Pamphlett *et al.*, 1995; Johnstone and Ambusaidi, 2000). Many universities face one major problem which is the large number of students in one course. Students need a lot of effort from their teachers in terms of teaching and assessment. Some assessment methods cannot realistically be used to assess large classes.

c. It is easy to develop banks of improved items based on statistical analysis of answers (Johnstone *et al.*, 1983; Pamphett *et al.*, 1995). Teachers are able to identify items on the test which were too easy, which were too hard and which were ambiguous. In addition, teachers can gather by pre-testing, a larger number of items than will be needed. The results from pre-testing will help teachers to discard those items, which seem unsuitable and modify others. Heywood (1977) lists the benefits of using pre-testing.

- Without pre-testing it is likely that some of the items will be badly designed and will not discriminate between students.
- Pre test/test data provide information about the reliability of both the questions and the examination as a whole.
- Pre-testing enables item writers to develop their skills.
- Pre-testing gives the pupils experience in answering this type of question.
- Pre-testing allows for experimentation in which pupil attitudes and difficulties can be evaluated alongside the teacher's view.

d. MCQs have a satisfactory test inter - marker reliability (Daniels, 1968; Tamir, 1990; Schmidt *et al.*, 1992; Verma *et al.*, 1997; Black, 1998; Staff Development and Educational Methods Unit). It is a very important feature of these types of questions since open response questions have a reliability problem. Open response questions have the 'halo' effect problem which could be described as the effect of the marker as a result of knowing the student or from previous test papers. This problem can be solved by using fixed response questions such as MCQs.

e. They are easily and quickly scored and lend themselves to machine scoring. This is one of the most attractive points to many teachers to use MCQs especially in higher education. Marking processes do not require any special skills from the marker and can be done by anyone even the students themselves (Currie, 1986a). Because of using machine marking, the scoring process makes it possible to return quickly and reduces the possibility of mistakes.

f. They can be seen as a suitable method of testing students who know their subject matter but are poor writers (Jackson 1988; Tamir, 1990; Child, 1997). Nitko (1983) proposes that these types of questions help students to focus on reading and thinking and so do not require the writing process to occur under examination conditions. Some students have a problem in the writing process and therefore fixed response questions such as MCQs can be seen as the best assessment method to assess their performance. The achievement of these students in MCQs does not depend on their writing skills.

g. They can be used for diagnostic purposes (Friel *et al.*, 1988; Treagust, 1988). If the distractors are based on common errors and misconceptions, then the items may give a diagnostic insight into difficulties an individual student may face. The results of such questions could be very helpful for a teacher to modify and correct his teaching methods. A recent study by Tan *et al.* (1999) found that multiple choice provides an easy - to - administer tool, providing results in a readily accessible form.

iv. Disadvantages of the Multiple Choice Questions:

Some educators, however, are against using MCQs in their assessment. They have their own evidence to support their views.

a. It is a less valid instrument compared to essay type questions (Verma *et al.*, 1997). The validity here means that the test is measuring what it was intended to measure in terms of knowledge, understanding and skills. In a multiple - choice test item some students may be able to eliminate several of the response options (correct and incorrect answers) because the options give unintended clues (Capper, 1996). The following are examples of invalidity of the MCQs.

Example One:

Select the word that means about the same as the word **SMART**.

a. clever. **b.** stupid. **c.** unfortunate. **d.** poor.

In this example, students who do not know the correct answer might be able to guess the answer because the connotations of the incorrect words are all very different from the correct answer. The words 'stupid', 'unfortunate' and 'poor' all have negative connotations and 'clever' stands out because it is the only option from the rest with positive connotations. If a student knows that the word 'smart' has positive connotation, he can then eliminate the three negative connotation options without knowing what is the meaning of the word 'smart'.

Example Two:

This is another example from science this time. Students are shown a picture of two girls standing side by side. From the picture, it is clear that these girls are of the same height. It also says that the girls are in the same class at school. The question is 'In which one of the following respects are the girls most likely to differ?'

a. height. **b.** age. **c.** rate of heart beat. **d.** number of toes.

The answer of the above question, by common sense, is option **c**. Students will eliminate options **a**, **b** and **d** to reach the correct answer. The correct answer could be reached by someone who never did science.

The **question** which could be pointed here is that "*Are we really measuring what we set out to measure in the multiple choice test?*". If the test is invalid then everything which follows from it is useless even if a good marking process or statistical manipulation is carried out. The test designer should be aware about the validity of the test he designs to ensure that the test does really measure what it is set up to do.

b. There is a problem of guessing and cueing effects (Handy *et al.*, 1973a;.Handy *et al.*, 1973b; Schawirth *et al.*, 1996). This is a major problem in educational and psychological measurement (Wang *et al.*, 1997). In four option MCQs the percentage of guessing is 25%. An ignorant pupil can score 25 % by random guessing. Some educators criticise the MCQs in terms that pupils can randomize their guessing. A study by Johnstone and his co - workers (Johnstone *et al.*, 1983) showed that the guessing could be increased from 25 % (four options) to 50 % because pupils can eliminate some distractors before guessing becomes necessary.

There are some suggestions which could be used to reduce guessing of MCQs:

- Increasing the number of options per item. It is less likely for the pupils to select the correct answer by chance. But this will bring us back the problem of finding more realistic options that could function well.
- Increasing the number of items per test. Again it is difficult to write good questions and plausible distractors for each question.
- Penalising or negative scoring of incorrectly answered questions. This may help to reduce guessing but, on the other hand, it may increase students' anxiety (Pamphlett *et al*, 1995). If the purpose of the test is to put students into an order of merit, there is

little point in penalising students for guessing. The rank order correlation between the 'raw' and 'penalised' scores is usually greater than 0.95 (Burton *et al.*, 2000).

Some teachers or examiners may believe that the guessing could be eliminated by applying any "correct for guessing" formulae such as the followings (Hudson, 1969; Handy *et al.*, 1973a):

a. Deducting marks for incorrect answers:

$$S = R - \frac{W}{N - 1}$$

Where S is 'true' score,

R is no. of correct answers,

W is no. of incorrect answers,

N. is no. of options in each item.

b. Adding marks for omitted items:

$$S' = R + \frac{O}{N}$$

Where S' is 'true' score,

R is no. of correct answers,

O is no. of items omitted,

N. is no. of options in each item.

However, it was found that even adjustment of final scores by one of the above formulae does not eliminate or minimize the effect of chance when answers have been guessed (Hudson, 1969). Some educators such as Johnstone *et al.* (1972) argued that it is better for a candidate to make an informed guess rather than to shelter behind 'no opinion'.

c. There are problems related to MCQs options. This is what is called positional response bias. Positional response bias as defined by Cronbach (1950) is one type of response set, as selecting one response position on multiple choice tests more often, regardless of item content. The psychological basis behind this is that some students read every option and discriminate carefully, whereas some students merely read through the item to find a plausible answer, mark it, and go to the next item.

If the options were superficial and ridiculous, then the students could arrive at the correct answer by a process of elimination (Child, 1997). Few researches have been done on the positional response bias in multiple choice tests because of the two types of problems (Fagley, 1987):

1. <u>Methodological problems</u>: such as

- Failure randomizes the position of keyed responses.
- Inclusion of unequal numbers of **a** -, **b** -, **c**- and **d** keyed items.
- Failure to consider the effects of visual cues provided by answer sheets.

2. <u>Conceptual problems</u>: such as

- Definitional differences that led to divergent views about what evidence would demonstrate a positional response bias.

However, some of the above problems can be solved nowadays using computer based assessment packages. These packages have the ability to randomize the keyed response and to eliminate the effects of visual cues.

There are two types of problems regarding the options. The first problem is related to the key answer whereas the second problem is related to the plausible distractor. Firstly, let us give a brief summary of some research findings related to the positional response bias of the key answer. A study by McNamara *et al.* (1945) concluded that a tendency exists among examinees to favour certain response positions over others. The difficulty of such questions is determined by the positions in which the correct answer has been assigned or located. They found that for five - choice items, those items having right answers in the fourth position are the most difficult; those with right answers in the second and third positions are the easiest; and those with right answers in the first and fifth positions are of equal moderate difficulty. On the other hand, Marcus (1963) found that MCQs are relatively free of position preferences. He found that the position of the plausible distractor (the distractor, which is selected most among other distractors in some pre – test) logically accounts for any response bias than does a position preferences.

Another study by Payne (1951) found that an alternative gained six (could be the right option) percentage points on average when placed among the top options (in **a** or **b** positions) and two percentage points when placed among the bottom options (in **d** or **e** positions) in comparison with same option placed near the middle of the answer option. The study by Blunch (1984) showed that the answer option receives different support when placed in different positions among other options.

Ace *et al.* (1973) revealed that the item difficulty of MCQs depends on the placement of the correct answer. Thus, for five options multiple choice questions, those with the correct option in the fourth position appeared to be more difficult than those with the correct option in the second or third position. They emphasised that the correct response position is probably an important determinant of item difficulty. Cizek's (1994) study suggested that item performance can be affected to a smaller or greater degree by changes in correct option placement.

There is also the question of the research findings on the positional response bias of the most plausible distractor (strong distractor). In MCQs, every distractor is supposed to take a part in distracting the students, otherwise it will be useless as an option in the item. A study by Friel (Friel, 1976; Friel *et al.*, 1979) with 12 - year old school pupils showed that a small editing in the item without changing one word or science data can bring about significant changes in the item's statistical factors, such as the facility value of the question.

These two problems above may be solved by presenting the options in different order in the hope of cancelling out the respondents' preferences for the various positions. There are two reasons why it may be desirable to reorder the options in MCQs (Cizek, 1994):

- To achieve a balanced key (equalizing the number of times each option is the correct response).
- Because of concerns about test anxiety, it is sometimes desirable to avoid the situation where one response is the correct answer for several contiguous items; hence options may be reordered.

However, not always is it possible to reorder the options of multiple choice questions. This is due to the fact that sometimes a stylistic change is desirable to place the options in some logically or aesthetically appealing manner (for example numbers in order). Therefore, the **question** which might be asked here is that "*Does reordering the options of the MCQs eliminate such bias*?". Research by Blunch (1984) showed that the position bias in the MCQs generally is not eliminated through the rotation of the options. Bresnock *et al.* (1989) supported this conclusion and pointed out that conclusions regarding randomization to removal personal position preferences are not convincing.

d. Failure to measure high level cognitive or psychological processes (Gronlund. 1971: Bray, 1986; Zeidner, 1987). MCQs tend to focus on the lower level aspects of knowledge such as recall and understanding while neglecting higher level aspects like analysis, synthesis and evaluation (Hager, 1994). They seem to test factual recall or the simple understanding of facts.

In medical courses for example, MCQs are widely used as one of the main methods for assessing students' performance. However, they have been criticised as being artificial and unrelated to clinical practice (Fenderson *et al.*, 1997). The way that these questions are presented to students may provide students with possible inappropriate cues.

Furthermore, it is difficult to test reasoning and problem - solving skills using MCQs because it is often not difficult for clever students to identify correct answer. MCQs may be able to measure what the pupil knows or understands but not what to do when confronted with problem situations. They cannot determine how pupils will react and perform in an actual situation.

If any institution wants to move beyond the knowledge and comprehension levels of the learning outcomes, then MCQs will be inadequate in judging performance and application of science concepts (Rasp, 1998).

e. Penalising the creative examinee (Zeidner, 1987). There is no place in MCQs for students to show their creativity. This is due to the fact that MCQs do not require from students any writing or open response. Research done by Frederiksen and his colleague (Frederiksen *et al.*, 1980) on measuring students' creativity, found that creativity cannot be measured with MCQs. If the creativity is defined as number of ideas, number of unusual ideas and number of ideas that are both unusual and of high quality, then the multiple choice method is not the appropriate method for measuring creativity.

f. Laying too much emphasis on speed and rote memory (Tamir, 1990). Currie (1986b) looks to examinations concerned mainly with recall of information such as MCQs as the test of speed - working against the clock. MCQs require no more than recognition of correct facts. Moreover, they encourage students to adopt surface approaches to studying (Entwistle *et al.*, 1992; Scouller, 1998). In most cases the way students are studying is determined by the nature of exam they have. If for example, the exam just requires students to select the correct answer from the given options, then students will opt for memorizing facts and no more.

Moreover, the type of questions are asked in any examinations influences not only the demands they make on students to answer them but also the form of understanding which students are seeking in their studying. Entwistle *et al.* (1991) set out to examine the links between the revision strategies adopted and the forms of understanding reached. They concluded that some students gear their revision to question types which can be answered within frameworks provided by the teacher or a textbook and that the type of questions set has a strong influence on the forms of understanding students seek during their revision study. They found that some types of questions encourage and test a restricted form of conceptual understanding. It could be reported that students are strongly influenced by the form of assessment they expect. Multiple choice formats push students towards surface approaches, while open essay type questions encourage a deep approach.

In addition, it has been suggested that the type of questions asked in the examinations can shape the content and the nature of instructions (Nitko, 1983). If the type of questions in the exam asks information in higher order skills, then the instructions will concentrate in that level and vice versa. It could be said that MCQs can shape education in undesirable ways.

g. MCQs seem to have language effect. Some researchers (Cassels *et al.*, 1984, Cassels *et al.*, 1985) found that small changes in the language of the item could affect the students performance in a test. In his study, Cassels *et al.*, (1984) found that there is:

1. A small effect in item difficulty by the use of active or passive voice.
2. Large effects occur when there are changes in:

- The key words in the item stems.
- Terms involving quantities.
- Overall complexity and length of the item stems.
- Using positive versus negatively worded items.
- Reducing thinking stages.

h. Students may make mistakes for a good reason or may get the right answer for the wrong reason. Sometimes it is difficult to find the reasons why some students made mistakes in MCQs. For example, if a student interprets a problem differently from the problem intended by the assessor, then what is the student going to do. In this case, there are two things that student may do; either leave the answer blank or make a guess. It is possible that the student failed to find any response which matched the solution he had in mind. A study by Tamir (1990) showed that up to a third of pupils who chose the correct response in MCQs might do it for a wrong reason.

i. It is difficult and time consuming to prepare MCQs (Stenhouse, 1969; Whitfield, 1970; Heywood, 1977; Child, 1997). This type of questions is good for sampling knowledge but it is much more difficult to construct them to test higher - order skills (Wiseman *et al.*, 1970; Currie, 1986a). The designer of MCQs should have a high level of experience in setting good questions. Verma *et al.* (1997) state that it is very easy to construct poor - quality MCQs and the results obtained from such MCQs tend to gain respectability under the guise of reliability.

The difficulty mainly comes from finding realistic options. If the options are not plausible then the chance of guessing the correct answer is increased (Bannister *et al.*, 1967; Schmidt *et al.*, 1992). The result of this is that the test distractors are non - functioning.

j. Cannot examine written expression (Head, 1966; Jackson, 1988). MCQs do not allow students to express their own ideas and show the linkage between their ideas because they required no more to do than choose among a fixed list of options. Complex structures of knowledge and reasoning cannot be assessed using MCQs because they will be in an isolated or restricted context (Black, 1998). Rowntree (1987) makes the thing very clear and condemns using one method, such as MCQs, in assessing students' performance. He states that example from a teaching and learning system that relied entirely on, say, multiple choice testing. Because of students' expectations of such tests, it might suffer the side effect of students learning to scan the learning experience for factual, testable items, gobbets of detail or technical terms, rather than looking for larger, encompassing insight that might be demanded of them in essays.

k. In MCQs some useful information is lost if no credit is gained for partial knowledge and if no analysis is being made to pupils' responses to different distractors (Friel *et al.*, 1978; Johnstone *et al.*, 1979). Most of the time the incorrect responses give very important information as a feedback to the tester about such test. Unfortunately, the current practice of the multiple choice tests is rewarded scores only for the correct responses. Therefore, it is recommended that instead of focusing on correct response only, incorrect responses to multiple choice items are used to extract more information about examinees' states of knowledge (Ndolichako *et al.*, 1997).

In addition, the scoring system of the traditional MCQs (one correct answer) does not take into account that sometimes each item is different from others. The weight that is given to each item is the same despite the fact that some items may contain more important information than others and vary in difficulty.

I. They cannot be used as criterion - referenced assessment. Criterion - referencing assessment refers to a means of measuring pupils' level of knowledge, skills, and understanding in the context of the task set (Baumann *et al.*, 1997). A driving test or a music exam are examples of criterion - referenced assessment.

There are two reasons why MCQs cannot be used as criterion - referencing assessment are listed by Johnstone *et al.* (1983).

- Criterion referencing is essentially checking the individual against a set of criteria. Objective item statistics refer to a population of individuals and tells nothing about the individual.
- Ideally, the facility value for a criterion - referenced item should be 1.0; that is, all the pupils have apparently succeeded in attaining a certain criterion. However, if an item has a facility value close to 1.0, it follows that the distractors are failing to distract.

Finally, we have to ask ourselves "*Can we avoid using MCQs in our assessment process?*". Some researchers such as Rasp (1998) still argue that we cannot avoid using MCQs in our testing. She points out that whether we like it or not, multiple choice tests are still a fact of life for our students and therefore, they need the practice.

2.2.1.2 True /False and Linked True/False:

i. Descriptions of True/False Items:

True/false items are one of the commonest and oldest type of fixed response questions. This type of question consist of a statement or proposition which the examinee must judge and mark as either true or false (Nitko, 1983).

<u>Examples of True/False from Biology:</u>

* **Directions:** Read each of the following statements: If the statement is true, circle the '**T**'. If the statement is false, circle the '**F**'.

T F **1.** The green colouring material in a plant leaf is called chlorophyll.

T F **2.** The corolla of a flower includes petals and sepals.

T F **3.** Photosynthesis is the process by which leaves make food for a plant.

<u>ii. Advantages of True/False Items:</u>

True/false items have several advantages in assessing students. These advantages are common in all the fixed response questions. Some of these are as follows (Macintosh, 1969; Gronlund, 1971; Collins *et al.*, 1976; Nitko, 1983):

a. They are relatively easy to write and construct. However, it is easy to construct poor true/ false items but to construct unambiguous true/false items require an extremely high degree of skill.

b. They can cover a wide range of content within a relatively short period of testing but in some areas of the curriculum it is difficult to construct good true/false items.

c. They can be scored easily and objectively.

<u>iii. Disadvantages of True/False Items</u>:

On the other hand, true/false items have obvious disadvantages. Some of these are as follows (Gronlund, 1968; Marshall *et al.*, 1971; Ebel, 1972; Thyne, 1974):

a. They can be answered correctly by blind guessing. True/false items have a fifty - fifty chance of guessing the correct response and therefore they encourage students to guess.

b. They may encourage students to study and accept only oversimplified statements of truth and factual details. In this case, true/false items encourage students to concentrate on remembering isolated factual details and rely heavily on rote learning.

Modifications have been made to true/false items to overcome the problems of both MCQs and traditional true/false. In these modifications, the traditional true/false items were compiled into batteries of true/false items to form linked true/false items (Figure 2.1).

The modifications were developed to overcome the drawbacks of MCQs because MCQs are in effect true/false applied each distractor and by using the linked true/false items, teachers can get a view of the pupil's grasp on an objective and a chain of reasoning is revealed (Johnstone *et al.*, 1983). In linked true/false, the true/false items should be interlinked in such a way that blind guessing would show up as gross inconsistencies in the responses (Gray, 1997). As a result of this, a consistent set of responses can give the examiner confidence in the results.

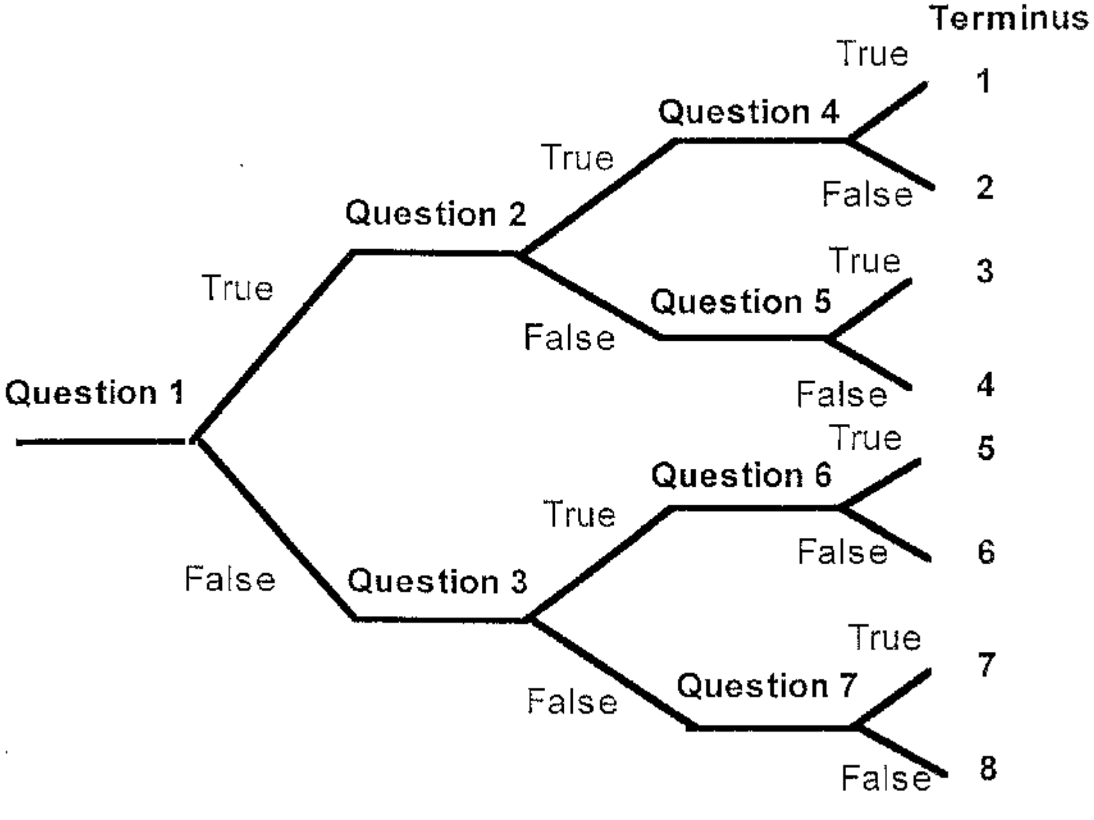


Figure 2.1: The layout of linked true/false questions (Adapted from Johnstone *et al.*, 1986)

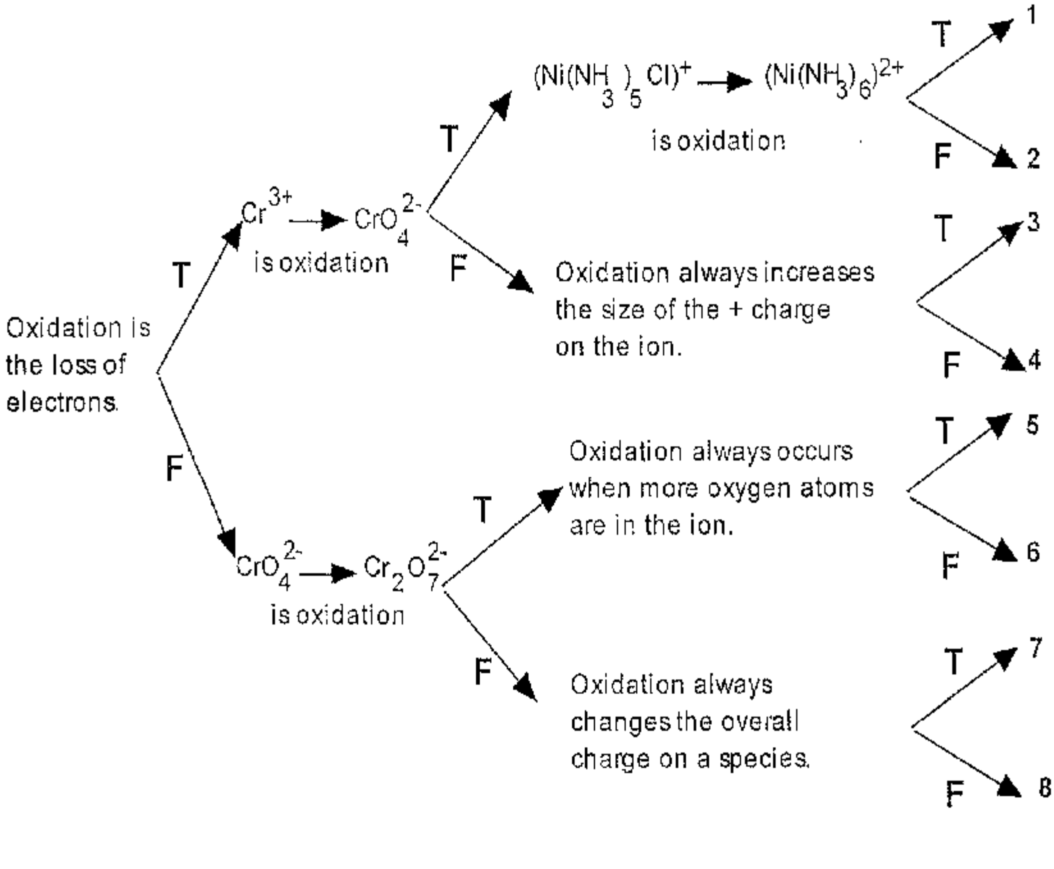
Figure 2.1 shows a layout of linked true/false items. The starting question (**Question 1**) is branched into two another questions (**Questions 2 and 3**) and each of these two questions is branched to another two until the end path has been reached. The student is required to answer each question by choosing true or false. Then, his answer leads him to another question and so on. Sometimes, the student may realize that he made mistake and chose the wrong path. In this case, he can go back and check his mistakes. The assessor may allocate marks only if the student reached the best end point or he may allocate marks for each correct answer.

From the diagnostic point of view, each path the student uses indicates the level of understanding the student has reached. Both the correct answers and wrong ones are so important in determining the level of understanding. For example, suppose a question consists of eight terminuses (like in Figure 2.1) and the best (correct) terminus is terminus number two. If the student reached this end point then, it can be said that this student reached the required level of understanding. However, if the student reached other than this terminus then, it is surely that this student has a partial or a complete lack of understanding of testing materials.

A computer program can be designed easily which could perform the testing, recording and analysis of results (Johnstone *et al.*, 1986). Unfortunately, linked true/false can be applied only in situations where black and white ideas are explored such as vertebrate/invertebrate, oxidation/reduction, metals/non - metals.

Example of Linked True/False from Chemistry:

* **Directions:** In the diagram below, decide whether each statement is True (**T**) or False (**F**) and follow that path.



2.2.1.3 Matching Questions:

i. Description of Matching Questions:

Matching questions are simply modifications of MCQs (Gronlund, 1968; Macintosh *et al.*, 1969; Marshall *et al.*, 1971; Thyne, 1974). In the MCQs the possible responses are listed underneath each individual stem, whereas in the matching questions a series of stems called ***premises*** is listed in one column and the ***responses*** are listed in another. Matching questions seem to be the appropriate method for identifying the relationships between two things (Gronlund, 1971).

Example of Matching Questions from Biology:

* **Directions:** You have two columns **A** and **B**. Column **A** contains a list of structures of the plant cell. On the line at the left of each statement, put the letter of the item in column **B** that best fits the statement. Each response in column **B** may be used once, more than once, or not at all.

Column A	**Column B**
____ 1. Nucleus	A. Gives plant cell definite shape.
____ 2. Vacuole	B. Carry out photosynthesis.
____ 3. Cell Wall	C. Occupy most of the cell.
____ 4. Cytoplasm	D. Site of the chemical reactions.
____ 5. Chloroplast	E. Controls the cell activities.
	F. Controls the movement of the substance in and from the cell.
	G. Able to stretch slightly when water is absorbed by the cell.

ii. Advantages of Matching Questions:

Although matching questions are a modification of the MCQs, there are some advantages in using them to assess students in science.

a. They can be scored rapidly, accurately and objectively by individuals who are unqualified to teach in the subject area being examined (Marshall *et al.*, 1971; Collins *et al.*, 1976).

b. There is a relatively low chance of guessing in matching questions if the number of responses is more than the number of premises. However, if there is an equal number of premises and responses, then the process of elimination improves the chance of guessing as the student works through the question (Collins *et al.*, 1976).

iii. Disadvantages of Matching Questions:

On the other hand, matching questions have limitations or disadvantages for assessing students.

a. It is difficult to prepare matching questions that can be used to test the higher levels of the cognitive domain (Gronlund, 1971; Marshall *et al.*, 1971; Collins *et al.*, 1976). This is one of the major problem of most of the traditional methods of fixed response questions.

b. If the premises are not developed in a homogeneous manner, they may provide clues to the correct responses (Marshall *et al.*, 1971). It is difficult to find enough important and homogeneous ideas to form a matching set. These types of questions cannot test a large sample of the course content.

c. Students may develop the undesirable study habit of memorizing isolated facts if they know that the teacher uses only matching questions (Collins *et al.*, 1976).

However, improvements have been made to this type of questions in United States and especially in medical schools. In this improvement, which is called Extended Matching Questions, students are asked to select the best answer to a question from a list of 20 options, each option may be used once, more than once, or not at all (Fenderson *et al.*, 1997). But is it easy to construct homogeneous premises to avoid cueing and guessing the correct answer?

2.2.1.4 Venn Diagrams Method:

i. Introduction:

This is another type of fixed response questions. It is the contribution of modern mathematics in science teaching (Johnson, 1977; Henson *et al.*, 1979). There is little research which has been done on this method and therefore, it is not in common use.

ii. Description of the Venn Diagrams:

In this method, students are presented with a diagram made up of intersecting circles (Figure 2.2). It could be two circles or more than that and usually each circle is labelled with a name. Then, students are asked to allocate the given information to an appropriate circle or overlap. It is not necessary that each piece of information has a place in the diagrams. Some of the information can be used as distractors to students. In this method, a teacher may allocate different number of marks for each correct placement of the information in the diagram. For example, in the diagram below (Figure 2.2) we assume that there is an object called **I** belongs to the three categories (**A**, **B** and **C**). If student places this object in ***Area 4***, then he will receive three marks because this object belongs to the three categories. However, if the same student places the same object in ***Areas*** *2, 5* and *6*, then he will receive two marks. The same student will receive only one mark if he places the same object in ***Areas*** *1, 3* and *7*.

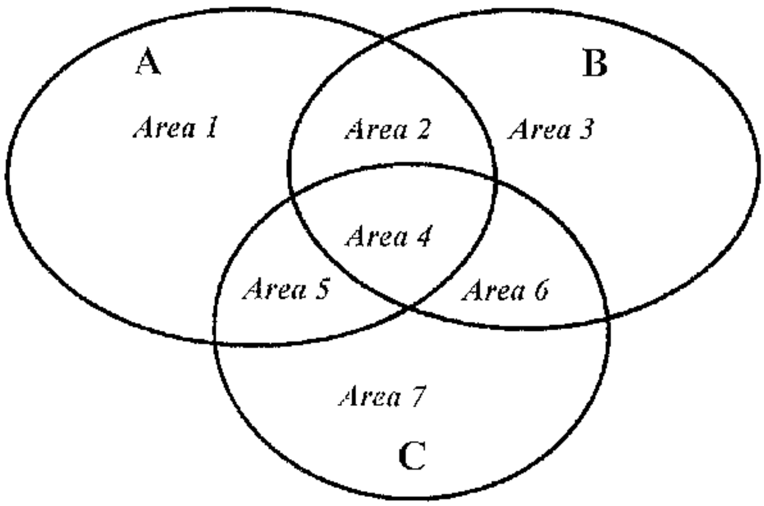


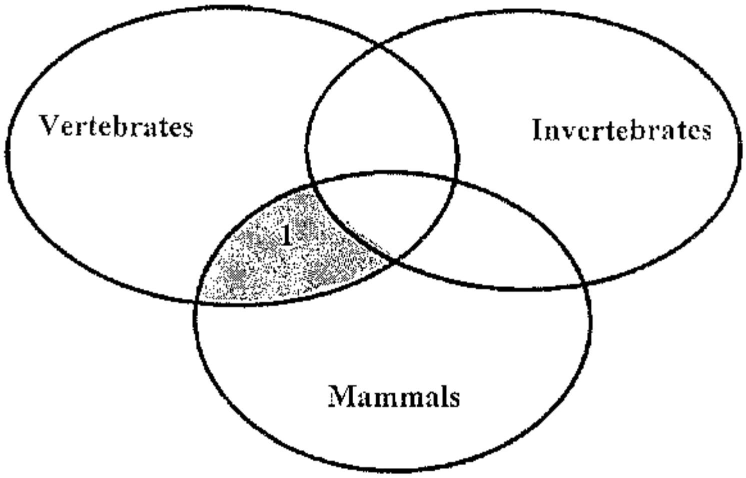
Figure 2.2: The layout of the Venn Diagrams method of assessment

The best way to understand any method of assessment is by giving an actual example. The following is an actual example of the Venn Diagrams from science.

Example of Venn Diagrams from Biology:

* Consider the following animals:

1. Coyote **2.** Red squirrel **3.** Trichinella spiralis **4.** Planarians **5.** Anteaters
6. Leeches **7.** Urodeles **8.** Alligators **9.** Sponge **10.** Manatee **11.** Lamprey



In the diagram above, animal number **1** (Coyote) has been placed in the shaded area because animal 1 is both vertebrate and mammal. Now write the numbers **2** to **11** in their appropriate places on the diagram.

There are other examples from science which can be assessed using Venn diagrams. For examples metal/non - metal, acid/base, and acidic oxide/alkaline oxide in chemistry and animal/plant, prokaryotes/eukaryotes, and meiosis/mitosis in biology.

iii. Advantages of Venn Diagrams:

Due to the fact that little research has been done in this method, the advantages and disadvantages of this method are not clear. However, a few things can be said about the advantages of the Venn diagrams method.

a. The big advantage of Venn diagrams method is testing for partial knowledge. For example, if Coyote (in the example above) was placed in **mammal** only, the student will get some credit, but his knowledge is not as well as developed as putting it in the **mammal / vertebrate overlap**.

b. It is a good method for examining the linkages between the concepts and investigates students' misconceptions. By asking students to label the given concept into a category may show how this concept links to other concepts. Also if one concept is labelled wrongly, a teacher can identify the misconceptions that students have in that particular concept(s).

c. They can be used to provide the teacher with the means of diagnostic assessment of the pupil's prior knowledge of specific concepts in science. This method can be used as a tool for diagnostic purposes before and after the teaching and learning process takes place. From the results, a teacher can build the starting point of his teaching and modify the teaching method(s) he uses.

d. Because it is one type of fixed response questions, it must be a reliable method and can be scored objectively. All fixed response questions avoid the subjectivity in marking that can occur with open response questions.

e. The visual presentation of Venn diagrams is simple, illuminating and easy to remember (Henson *et al.*, 1979).

***iv. Disadvantages of Venn Diagrams*:**

The disadvantage of this method which could be identified is that it may only be suitable for some parts of the curriculum that show some kind of linkages between the concepts. In other words, this method is only suitable for testing the skill of categorisation.

After the above discussions of some fixed response methods, the **key question** is "*Is there any other fixed response method that could overcome the problems which the above methods have?*" The answer to this question is "***Yes there is***". This method is called structural communication grids and the current research is intended to investigate it. However, it is important first to give a background review about this method by looking to what has been written in the literature. An attempt has been made to investigate the effect of some factors on students' responses in this method.

2.2.1.5 Structural Communication Grids (SCGs):

i. Introduction:

The are two reasons why there is a need for finding new assessment methods in assessing students' performance in science curriculum. The first one is related to the internal and external pressure mentioned in chapter one. This pressure is due to the following reasons:

- Much assessment of the students is carried out, consciously or unconsciously, in so many different ways and with so many different purposes (Jackson, 1988). This assessment has a lot of weight and affects the futures of the students. Therefore, effective assessment methods are needed.
- The big movement in the learning and teaching process is resulting in new methods of teaching. These methods such as, problem solving, help to promote effective learning and teaching. The assessment process should be paralleled to any development and improvement in teaching and learning.
- The changing aims of higher education such as mix of academic, vocational and general transferable skills.
- Emphasizing deep learning rather than surface (shallow).
- To save staff time.

The second one is related to the multiple choice questions (MCQs). MCQs have many disadvantages when used as the main method for assessing students' performance in science. As a result of this, science educators have to think of different methods that can match the requirement of the internal and external pressure and overcome the problem of the MCQs. One of the new methods which has been used to assess students' performance in science is what is called structural communication grids (SCGs).

ii. Description of the Structural Communication Grids:

SCGs can be classified as one type of objective testing because there is no chance for subjectivity in marking such questions. They were invented and developed in England (Egan, 1972; Egan, 1973). They are not well known to teachers and researchers because of the lack of the literature. However, most of the research on SCGs has been carried out in the Centre for Science Education, University of Glasgow. The reason for SCGs appearing is the increase in the quantity of students in schools and universities that affect teacher/students ratio. The loss which is becoming more apparent in increasing numbers of students is in depth of understanding of the curricular materials learned, and thus also a loss in the ability to use what is learned flexibility and creativity. The pressures of examinations, which can usually be passed on the strength of the "stored" information tend to exacerbate the problem (Egan, 1972). SCGs can be used for different purposes such as a teaching and learning tool and an assessment method. Nowadays these questions are quite widely used by the Scottish Qualifications Authority to assess pupils' performance in science at both Standard and Higher - Grade and mainly in chemistry. At the university level, these types of questions are not common and it is hoped to see these questions in the university examinations in the future.

Egan described SCGs as a means by which a pupil constructs his response to a problem by selecting a number of items from the matrix provided (Egan, 1976; MacGuire *et al.*, 1987). In other words, pupils are given a grid or a matrix with different concepts or statements from which a selection is made to answer a battery of questions. Different problems have been asked and students are required to answer these problems based on the concepts or statements in the grid. In each question or problem, the knowledge elements are 'structured' in such away that, as the student builds his response, the various patterns and relationships of meaning emerge (Egan, 1973).

The aim of the SCGs is to present the students in each question with a challenge asking him/her to organise an undetermined number of items in the grid into a coherent and satisfactory solution (Egan, 1972; Egan, 1976). In each grid, both the relevant and non - relevant information for a particular question exist.

The differences between SCGs and MCQs is that in SCGs the student must select as many pieces of information as he thinks are relevant to answer the question. He must then order his selection in a logical way to answer the question. Moreover, in SCGs, the student does not know in advance how many pieces are needed to answer a question like in multiple choice questions. The guessing is almost eliminated in this type of question. SCGs differ from the essay questions in a way that it overcomes the problem of tackling the unmanageable subjective elements that essay questions have.

The appropriate size of the grid depends on the age of the students. At school level, grids with six boxes are used. However, at university level the number of boxes in the grid is larger. For example, Johnstone (1988) used grid questions successfully for first year students with twelve boxes either 4 × 3 or 3 × 4. For fourth year students grid questions with sixteen boxes have been used. Any examiner should consider the age of his examinee before setting up any grid questions. It is likely that older examinee can cope with larger grid.

A Simple Example of SCGs from Chemistry:

Many processes take place in the petrochemical industry.

blending A	hydration B	reforming C
polymerisation D	cracking E	oxidation F

Which box(es) contain(s) the process(es) taking place in the production of poly(ethene) from naphtha?

A	B	C
D	E	F

In the above question, pupil needs to draw a circle around the appropriate letter. The assessor can ask more than one question with one or more common options. Therefore, pupils cannot eliminate any option because it could be used again.

However, the use of the grid questions can go beyond asking direct questions. The assessor can ask pupils to select the relevant information from irrelevant ones and to put the relevant ones in a logical order. The logical order takes various forms such as the order in terms of increasing or decreasing of something or in terms of the time the events happening or the size of something or just a logical sequence of ideas.

Example of Selection and Ordering in SCGs from Biology:

The grid below contains the metabolic reactions which happen during mobilisation of food stores in seed germination. Use the numbers from the boxes to answer the following questions. Each number can be used once or more than once.

Activation of embryo 1	Synthesis of Alpha Amylase 2	Breakdown of Starch to Glucose 3
Releasing of Gibberelins 4	Synthesis of a Protease 5	Diffusion of Glucose or Amino Acid to embryo 6
Stimulation of Aleurone Layer 7	Water absorption of seed 8	Breakdown of Protein to Amino Acids 9

Q1. Which boxes contain the reactions which happen during mobilisation of carbohydrate reserves in seed germination?

a. Select the relevant boxes?..

b. Put your selection into a logical sequence?.................................

(Relevant boxes and their sequence: **8, 1, 4, 7, 2, 3, 6**)

Q2. Which boxes contain the reactions which happen during mobilisation of protein reserves in seed germination?

a. Select the relevant boxes?..

b. Put your selection into a logical sequence?.................................

(Relevant boxes and their sequence: **8, 1, 4, 7, 5, 9, 6**)

iii. Methods of Scoring Structural Communication Grids:

This section aims to give a summary of scoring methods which have been used to score the selection and the logical order in grid questions. The first part will discuss how to score the selection and the second part will deal with scoring the sequence ordering of the selection.

a. Scoring the Selection:

There are different methods that could be used to score the selection in grid questions. However, only two methods will be discussed in this section as they seem to be more logical, fair and are in use by some institutions.

a.1 Method One: Scottish Qualifications Authority:

In the Scottish Qualifications Authority scoring method, the grid questions are divided into two types; closed and open questions. Closed questions are those questions in which the assessor indicates to the pupils in the stem of the question how many selection(s) are required to answer the question. One mark has been allocated for this type of question despite the number of correct selection(s). A pupil should select only the correct selection(s) in order to receive one mark. A pupil will be penalized if more or less than what is required is selected even if one of his selections is correct. In terms of wrong selection(s), there is no mark allocated. Closed questions are easier because they do not require a lot of thinking from pupils. Furthermore, guessing in this type is more likely to happen than in open ones.

The second type of question (i.e. open questions) are called open because the assessor does not indicate in the stem how many correct selection(s) that pupil should make. One mark has been allocated for each correct selection. If there are two correct answers then the pupil will receive two marks and so on. In terms of selecting more than required, he will be penalised by cutting one mark from his scores (Table 2.1).

Table 2.1: Number of marks allocated for each type of selection(s) in closed and open questions of the Scottish Qualifications Authority scoring method

	Types of Questions	
	Closed	**Open**
Marks allocated for the correct selection(s)	one mark for whole question	one mark for each correct selection
Select more than required	no marks	subtract one mark
Select less than required	no marks	no mark subtracted
Select two, one correct and one wrong if two correct selections are required	no marks	one mark
All selections wrong	no marks	no marks

Now let us take an actual example of the above method of scoring to make it easy to understand.

Example of Scottish Qualifications Authority Scoring Method from Chemistry:

* Many processes take place in the petrochemical industry.

blending **A**	hydration **B**	reforming **C**
polymerisation **D**	cracking **E**	oxidation **F**

Q1. Which box contains the process taking place in the production of ethanol from ethene?

A	B	C
D	E	F

Q2. Which box(es) contain(s) the process(es) taking place in the production of poly(ethene) from naphtha?

A	B	C
D	E	F

The correct answer for question (1) is box **B** (hydration) and for question (2) are boxes **D** and **E** (polymerisation and cracking). In the example, question 1 is a closed type question whereas question **2** is an open type question. If pupil **X,** for example, answered question 1 correctly, then she will be allocated one mark. However, no mark will be allocated if she got it wrong or selected two options; one correct and one wrong.

For question (2) the same pupil will be allocated two marks if she selected the two correct answers. But if this pupil selected one correct answer and one wrong answer, then she will get only one mark for the correct one and no mark for the incorrect. If she selected three answers; two corrects and one wrong, then she will receive only one mark for the question. Finally, if the same pupil did not select any correct answer(s), then she will not gain any mark.

At the school level, the grid questions which are used to assess pupils are dealing only with the selection part. There is no place for putting pupils' selections in a logical order.

a.2 Method Two: Method Suggested by Egan:

This method is quite difficult to do by hand, but it is much easier if it is done by computer or from a pre-calculated table. There is a penalty for wrong selection. Therefore, a pupil has to think before answering the questions otherwise he will be penalized if he makes a mistake by guessing.

The following steps show how to score the selection(s) of the relevant information from the grid using Egan's method.

Step one: Calculate the "coefficient of confusion" using the formula below. Egan coined the term "coefficient of confusion" as a mean of scoring.

Coefficient of Confusion (Scores) =

$$\frac{\text{No. of relevant } (\checkmark) \text{ responses chosen}}{\text{No. of relevant } (\checkmark) \text{ responses available}} - \frac{\text{No. of irrelevant } (\times) \text{ responses chosen}}{\text{No. of irrelevant } (\times) \text{ responses available}}$$

The resultant value of the "coefficient of confusion" range from +1, through 0 to -1.

Step two: The assessor then may decide to multiply or add to the value of the coefficient by any factor to reflect the importance of the item.

a. In terms of multiplication, either by the number of correct answer chosen or any other factor such as **5** or **10**. In the multiplication, there is a possibility that a pupil will get negative scores.

b. However, the assessor may decide to add (+1) to the "coefficient of confusion" to eliminate the negative. Here the range value will be then between **0** and **+ 2.** Then this can be multiplied by a factor to reflect the 'value' of the question.

Now let us take an actual example of Egan's method of scoring the selection. We will take the example mentioned in page 30. The correct selection of question **1**, in this example, is boxes **1, 2, 3, 4, 6, 7, 8**. Suppose that the following students made their selections as follows:

Student (A) = 1, 2, 3, 4, 6, 7, 8.

Student (B) = 1, 2, 3, 4, 6, 7, 9.

Student (C) = 1, 2, 4, 5, 6, 8, 9.

Student (D) = 2, 3, 4, 6, 7, 8.

Student (E) = 2, 3, 5, 6, 7, 8.

Student (F) = 1, 2, 3, 4, 5, 9.

Student (G) = 2, 3, 4, 6, 7.

Student (H) = 2, 3, 5, 7, 8.

Student (I) = 1, 5, 6, 8, 9.

Student (J) = 3, 4, 6, 7.

Student (K) = 3, 4, 5, 6.

Student (L) = 1, 2, 5, 9.

Student (M) = 3, 4, 6.

Student (N) = 2, 3, 9.

Student (O) = 1, 5, 9.

Student (P) = 1, 5.

Student (Q) = 5, 9.

The scores allocated for each of the above students can be found in the Table 2.2 below.

Table 2.2: The marks allocated for some students according to their selection using "coefficient of confusion" formula

Student	No. of Selections	Type of Selection	Coefficient of Confusion (Marks)
A	7	Seven Correct	1
B	7	Six Correct and One Wrong	0.36
C	7	Five Correct and Two Wrongs	-0.29
D	6	Six Correct	0.85
E	6	Five Correct and One Wrong	0.21
F	6	Four Correct and Two Wrongs	- 0.43
G	5	Five Correct	0.71
H	5	Four Correct and One Wrong	0.07
I	5	Three Correct and Two Wrongs	- 0.57
J	4	Four Correct	0.57
K	4	Three Correct and One Wrong	- 0.07
L	4	Two Correct and Two Wrongs	- 0.71
M	3	Three Correct	0.43
N	3	Two Correct and One Wrong	- 0.21
O	3	One Correct and Two Wrongs	- 0.86
P	2	One Correct and One Wrong	- 0.36
Q	2	Two Wrongs	-1

However, the assessor should be very careful when he writes and constructs the test based on SCGs and scores them by "**coefficient of confusion**" formula. The number of answers should not exceed half (1/2) of total number of boxes available, otherwise he will give unfair marks to his students. Student (F), in the above example, has made a better selection (four correct and two wrongs) compared to student (**P**) (one correct and one wrong), but student (**P**) received better scores than student (**F**) (Scores for student (**F**) = - 0.43, whereas for student (**P**) = - 0.36). The same problem appeared between students (**C**) and (**N**). Although student (**C**) has better selection (five correct and two wrongs) than (**N**) (two correct and one wrong), student (**N**) received better scores than the student (**C**) (Scores for student (**C**) = - 0.29, whereas for student (**N**) = - 0.21).

After giving example of both scoring methods of selection part of the SCGs, it is worth giving a summary of the two methods (Table 2.3).

Table 2.3: Comparison between Scottish Qualifications Authority Scoring Method (Method One) and Method Suggested by Egan (Method Two) of scoring the selection part of the SCGs

	Method One	Method Two
Type of Questions	Questions are divided into closed and open.	Questions are not divided.
Partial Marks	No partial marks in the closed questions but there is in the open one.	There is a partial mark which depends on the selection(s) made by the pupil.
Fairness	Less fair because pupil will not receive any marks if he selected more or less than required in the closed questions.	Fairer. Pupil will receive marks on his selection depending on the quality of the selection(s) he made.
Negative Marking	No negative marking	There is negative marking but the assessor can eliminate this by adding +1 to the value of "coefficient of confusion".

b. Scoring the Logical Order of the Selection:

The big problem facing the grid questions is how to score the ordering part of the selection if it is required in the question.

There is a suggestion made by the designer of the computer based assessment package which is known as Tripartite Interactive Assessment Delivery System (**TRIADS**). Mackenzie (1997) suggested three methods of scoring the ordering of the pupils' selections.

Method One (Exact Match): In this method the answer sequence should be in exact match with the ideal sequence. The examinee will get either the full marks or zero. It is a very hard method but it is very useful for the safety of assessment.

Method Two (Before): In this method, the scores will be allocated to student only when he puts the statement **A** is before statement **B** and statement **B** is before statement **C** and statement **C** is before statement **D** and so on; no matter where they are in the sequence. Scores are only allocated for pairs of items if they exist in the manner described above. An example is the best way to show how this method is working. We assume that the ideal sequence of a selection in a question is boxes **1, 3, 5, 6, 8, 4, 2**. Student **X** has the following sequence **1, 5, 3, 6, 4, 8, 2**. This student will receive eight marks out of twelve. The eight marks are coming from the following:

⇒ 2 marks because statement **1** came before statement **3**.

⇒ no mark because statement **3** does not come before statement **5**.

⇒ 2 marks because statement **5** came before statement **6**.

⇒ 2 marks because statement 6 came before statement **8**.

⇒ no mark because statement **8** does not come before statement 4.

⇒ 2 marks because statement **4** came before statement **2**.

From the researcher's point of view, this method does not really reflect the linkages of the ideas in the student's mind. In addition, what about the first and the last statements if they are placed in the wrong positions. Consequently, there should be another method which may overcome the limitations existing in the above two methods.

Method Three (Before and Next To): This method looks more realistic and fair compared to the other two methods. The marks will be allocated only if the statement **A** comes before and next to (or adjacent to) statement **B**, statement **B** is before and next to statement **C** and statement **C** is before and next to statement **D** and so on. Scores are allocated for each pair of items in the ideal sequence. In this method, the assessor may want to divide the two marks between the **before** condition and **the next to or adjacent** condition. For example, suppose that the ideal sequence of a selection in a question is boxes **1, 3, 5, 6, 8, 4, 2**. One student has the following sequence **1, 3, 6, 5, 8, 4, 2**. This student will receive six marks out of twelve. The six marks are coming from the following:

⇒ 2 marks because statement **1** came before and next to statement **3**.

⇒ no mark because statement **3** came before but not next to statement **5**.

⇒ no mark because statement **5** does not come before statement **6**.

⇒ no marks because statement 6 came before but not next to statement **8**.

⇒ 2 marks because statement **8** came before and next to statement **4**.

⇒ 2 marks because statement **4** came before and next to statement **2**.

Let us now look to the following example as an application from science to the three methods of scoring the logical order.

Example of Scoring the Logical Order from Biology:

The grid below contains the metabolic reactions which happen during mobilisation of food stores in seed germination. Use the numbers from the boxes to answer the following questions. Each number can be used once or more than once.

Activation of embryo **1**	Synthesis of Alpha Amylase **2**	Breakdown of Starch to Glucose **3**
Releasing of Gibberelins **4**	Synthesis of a Protease **5**	Diffusion of Glucose or Amino Acid to embryo **6**
Stimulation of Aleurone Layer **7**	Water absorption of seed **8**	Breakdown of Protein to Amino Acids **9**

Q1. Which boxes contain the reactions which happen during mobilisation of carbohydrate reserves in seed germination?

a. Select the relevant boxes?...

b. Put your selection into a logical sequence?.................................

(Relevant boxes: **1, 2, 3, 4, 6, 7, 8**)

The ideal sequence of the above question is boxes **8, 1, 4, 7, 2, 3, 6**. We took four students who answered the sequence part as an example.

* Student **A** has the following sequence: **8, 1, 4, 7, 2, 3, 6.**

* Student **B** has the following sequence: **8, 4, 7, 1, 3, 2, 6.**

* Student **C** has the following sequence: **8, 3, 1, 7, 5, 4, 6.**

* Student **D** has the following sequence: **8, 1, 3, 7, 4, 2, 6.**

To know the number of marks allocated for each individual student above, see Table 2.4 below.

Table 2.4: Marks allocated for each of the four students according to their order of the selection in the example above

	Marks		
Student	**Method One (Exact Match)**	**Method Two (Before)**	**Method Three (Before and Next To)**
A	12	12	12
B	0	8	2
C	0	6	0
D	0	8	2

From the researcher's point of view, Mackenzie's third method **(Before and Next to)** is the most convenient method of scoring the logical order of the students' selections. This is because although Mackenzie's method two gives students more marks but it does not show exactly the logical order and the linkage of the ideas in the students' minds. Mackenzie's method one **(Exact Match)** is very hard and it is difficult to apply if the student has to order many pieces of information logically.

iv. The Use of Structural Communication Grids:

SCGs can be used for various purposes. Johnstone (1988) lists the use of grid questions as follow:

a. Testing the ability to categorise and dig into concepts: the examiner can use grid questions to ask students to categorise the given concepts. Moreover, he can use the technique to determine the misconceptions and misunderstanding a student has of the given concepts.

b. Testing the ability to sequence ideas: the examiner can ask students to sequence their selections logically. This will help an examiner to determine the problem that students have in any process. For example, the examiner might discover that the students did not understand part of Kreb's Cycle in biology when she asked them to order the steps of the cycle logically.

c. Testing descriptions or procedures: the examiner may ask students to select the boxes that describe something. For example, to select the boxes to describe the mitosis process in biology or formation of the solid sodium chloride from solid sodium and gaseous chlorine in chemistry.

d. Testing deductions and inferences at various levels: here the examiner gives students a variety of statements in the grid. Then, the student can put them in different ways which show the degree of understanding and level of the complexity of the deductions and inferences. The examiner may say to students that he has an idea in his mind and certain boxes are examples of that idea and certain boxes are not. Then, he asks students to give other examples of that idea. For example, the examiner has an idea about **Phylum Mollusca** in biology and *snails*, *slugs* and *clams* are examples of this phylum whereas *Trichinella spiralis* is not an example, then students have to recognise the common factor(s) and use them to extend the selection.

v. Advantages of Structural Communication Grids:

Although little research has been done in SCGs, it seems that using such technique has several advantages. The following points highlight these advantages (Egan, 1973; Duncan, 1974; Johnstone *et al.*, 1979; Johnstone *et al.*, 1983; Gallacher, 1984; MacGuire *et al.*, 1987; Johnstone, 1988; Johnstone *et al.*, 2000).

a. Can be used to provide the teacher with the means of diagnostic assessment of the pupil's prior knowledge of specific concepts in science. Gallacher (1984) suggested that structural communication is a useful alternative to the traditional methods of diagnostic assessment. A teacher may design a few grid questions in a topic (e.g. oxidation-reduction) to provide him with the knowledge that his pupils have about that topic. Moreover, the grid questions can provide the teacher with the existing type of understanding (partial or complete) in a pupil's mind. If the pupils have only partial understanding of the topic, then the teacher could base the starting point of his teaching of that topic based on this understanding. One of the recent studies on SCGs as a tool of diagnostic assessment was carried out by Bahar (1999). He used grid questions to examine pupils' misconceptions in some biological topics such as food digestion and the chemistry of respiration at secondary and tertiary levels.

b. Allowing an objective measure of the higher level intellectual skills. One of the criticisms of the fixed response methods of assessment, as it was mentioned previously, is their lack of testing the high order skills. These skills have been assessed by essay questions which have problems of reliability and objectivity. Structural communication grids can replace the essay questions to measure these skills objectively and score them in a reliable way. SCGs demand two of the same skills as an essay: the ability to separate relevant information from irrelevant and to present it in a logical way.

c. It is objectively markable with a highly reliable technique of testing. A study by Gallacher (1984) showed that grid questions are reliable even if the grids are presented in different shapes (e.g. 3×4 or 4×3) and when the information is arranged in a variety of ways across the grids. He concluded that this type of questions could be used to extend and increase the objectivity of extended answer questions. This is due to the nature of the grid questions which do not require any writing or expressing ideas by writing. Therefore, there is no subjectivity in the marking process. The problem of reliability and objectivity that some assessment methods possess can be overcome in grid questions.

d. It is one way of looking to the concepts that students have in their minds. The examiner has a mine of information about student's knowledge. From the pupil's selection, the examiner is able to know what a pupil has correctly chosen and what has been wrongly chosen. Also the examiner can determine the correct ideas which have been omitted. Furthermore, by asking pupils to place their selections in logical order, the teacher is able to identify the sort of linkage that each pupil has in his mind. In addition, the teacher is able to discover which concept(s) were not understood.

e. Can be used instead of the essay type questions. The major problem facing essay type questions is the lack of objectivity and the reliability of marking. As was mentioned in point (b) above, SCGs are highly reliable and objective in marking. An assessor may use it as an alternative method to essay questions. Essay questions require students to write and express their ideas in a paragraph. The examiner can put the statements of such a paragraph in a grid form and ask students to decide and select the information that he wishes to include in his paragraph and then construct a coherent unit from this information.

f. The contents of the boxes can be words, phrases, pictures, equations, definitions, numbers, formulae and so on. In other words, the contents of the boxes can be varied, so that they can be made suitable for visual as well as verbal thinkers.

vi. Disadvantages of Structural Communication Grids:

The researcher could not find any researches done on the disadvantages of SCGs. It is important to note that more research is needed to identify the limitations and problems of SCGs. However, one major aim of this research is to investigate and examine if there are any limitations in the use of grid questions in assessing students' performance in science.

Nevertheless, one point which should be mentioned under the heading of the disadvantages of SCGs is that of the size of the grid (i.e. the number of boxes in the grid). A lot of mental effort is involved in choosing the relevant responses from a large grid (Duncan, 1974). Therefore, the teacher must use the appropriate size of the grid that matches the level of pupils' study.

2.2.2 Open Response Questions:

2.2.2.1 Introduction:

Open response questions (essay questions) are another type of question widely used to assess students' performance in a science curriculum. Essay questions and MCQs have become the two most popular formats of questions employed in the construction of classroom achievement tests (Zeidner, 1987). The emphasis on using open response questions arises from the view that not everything in science can be assessed by fixed response questions.

2.2.2.2 Advantages of Open Response Questions:

There are many advantages and benefits of using open response questions in assessing students' performance (Gronlund, 1971; Entwistle *et al.*, 1991; Entwistle *et al.*, 1992; Knutton, 1994; Aiken, 1997; Verma *et al.*, 1997; Black, 1998).

a. Easy to set and prepare. This is one of the most attractive points of the open response questions to many assessors. These questions, as opposed to MCQs, do not require a lot of time to set and prepare. The assessor is not required to find plausible distractors to avoid pupils' guessing.

b. They encourage a deep approach to understanding during pupils' studying and revising. If pupils expect open response questions in their exam, they will adopt a deeper approach to understanding. Open response questions are not those questions which ask pupils to select from the given alternatives, but are questions in which the pupil has to express his ideas and show his abilities.

c. One criticism of MCQs is their inability to assess higher order skills. This limitation can be overcome by using open response questions. The examiner can use open response questions to measure the higher order skills such as synthesis, analysis and evaluation.

d. They require complex structures of knowledge and reasoning to be explored and there is an opportunity that pupils can express their own ideas. MCQs penalise the creative pupils because they do not give these pupils a chance to show their creativity and express their ideas. Open response questions, by contrast, give pupils space and opportunity to show their creativity and explore the pupil's capacity to select, assemble and integrate various facts of knowledge and understanding and to explain, and evaluate.

e. There is no chance of guessing. In open response questions, pupils need to recall and supply answers rather than recognise already supplied answers. However, the guessing comes in question spotting. Since extended answer questions are few in number, students can guess the topic that will come and learn these only.

f. They may be more valid instrument than MCQs in assessing what is intended to measure. As it was mentioned in section 2.2.1.1 that MCQs have the problem of the validity; open response questions seem to have better validity because they require from students to write and express their ideas. Moreover, it is a valid method of assessing pupils' higher order skills and abilities.

2.2.2.3 Disadvantages of Open Response Questions:

Open response questions, like any type of assessment format, have some limitations and disadvantages. It has been emphasised again that it is important that the assessor determines the intended outcomes to be assessed before choosing the assessment methods.

The following points summarise some limitations of using open response questions (Stenhouse, 1963; Gronlund, 1971; Chapman, 1973; Bray, 1986; Lorsbach *et al.*, 1992; Capper, 1996; Child, 1997; Black, 1998).

a. They are a less reliable type of question. One of the criticisms of open response questions is the subjectivity when marking. It is difficult to find two markers who give the same marks to one student in essay questions. It is argued that the reliability coefficient of essay questions can reach 0.30, which is very low compared to the reliability coefficient of MCQs which should be 1.0 (University of London, 1976). The problems of presentation and handwriting may play a part in marking pupils' essay questions subjectively. However, marking criteria (schemes), set up before the marking process takes place, may help to reduce the problem of subjectivity of open response questions.

b. It is very difficult to cover a substantial portion of a syllabus using essay type questions. One feature of open response questions is pupils have to write sentences in order to answer the questions. Therefore, it is not easy to ask pupils many open response questions in two hours exam.

c. There is a danger that student can write for an hour about something without necessarily saying anything sensible or getting his facts correct or covering properly even the main points of the topic and without eventually gaining marks. Nonetheless, this can be reduced if the given questions are highly structured and clear.

d. Sometimes a student knows the answer to the given questions but he cannot present the ideas in a coherent manner. Thus, a student loses marks although he knows the correct answer.

e. Amount of time requiring for scoring the answer. It takes a long time to score open response questions even if the scoring guidelines are well defined. This is why many teachers favour MCQs to assess students in higher education. If one teacher has more than 600 students in one course, he needs many hours to mark a test based on essay questions.

2.3 Computer Based Assessment:

2.3.1 Introduction:

There is a tendency these days towards using information technology (IT) in every single part of life. The **key question** which should be asked here is, "*Is there any role for information technology (i.e. computer) in students' assessment in science?*".

The reasons for including this section in the theoretical part of the thesis are because: firstly, part of the research was administered to students using computer based assessment packages. Secondly, a computer program for SCGs was specially designed to enable pupils to use them for self - assessment and to facilitate marking grid questions by encouraging teachers to use them.

Computers and audiovisual media have become widely used at many levels of education as a whole and especially in science instruction (Gidding *et al.*, 1991). There have been enormous ranges of methods, approaches and techniques designed to improve science instruction which are supposed to help students learn science better.

In the area of assessment, most use of the computer has been for statistical and scoring purposes. However, more recently there have been movements to use computers as a tool for delivering assessment to students, mainly for self - assessment and formative purposes. New software has been produced and designed for the purpose of using the computer as a tool for assessment, such as **Question Mark (QM)** for Windows and Tripartite Interactive Assessment Delivery System (**TRIADS**).

At the beginning, such packages dealt with questions mainly in the multiple - choice format. However, as the progress of developing such packages becomes more effective, other methods of assessment may be introduced.

2.3.2 Advantages of Computer Based Assessment:

Using the computer as a tool for assessment has many advantages (Jackson, 1988; Leonard *et al.*, 1997; Bowen, 1997).

a. Students like rapid feedback. This is one of the greater advantages of using information technology as a tool for assessment. The computer has the ability to mark the questions immediately after the students have answered the test. It displays the scores to students so that they know their answer is right or wrong. Students are able to identify their strengths and weaknesses in the course. The immediate feedback is a crucial part of the learning process.

b. A useful management tool (either for constructing or administering test) for keeping track of students' grades. It enables information to be presented in different ways to meet the needs of different audiences such as teachers, students...etc.

c. A useful method to know how students were achieving in the learning objectives of the course. In higher education computer based assessment becomes very important because of the increase number of students in one course and the consequent increase in the time spent by staff on assessment.

The study by Leonard and her colleagues (Leonard *et al.*, 1997) about the first year biology course at the University of Glasgow, where the number of students in one class was over 550, identified a degree of concern about knowing how students were achieving learning objectives of the course as it is progressed.

d. Self - assessment in the students' own time, at their own pace, when they are ready. Students can use the computer to practise similar types of questions they have at the end of the year. By doing this students become more familiar with the type of questions they are likely to meet in the formal exams. Furthermore, at the university level, it is not necessary that each student has the same timetables. Therefore, computer based assessment for self - assessment purposes allows students to practise at any time which suits them.

e. Increase students' confidence. Computer based assessment has a flexibility that allows students to practise more than once. This will help them to be more confident in answering questions. Also more practice of such questions helps students to be more confident in the subject. At the end, students become more confident with the format of the final exam questions.

f. Staff can be alerted sooner to adapt their teaching. The computer has the advantages of keeping a record of the students' responses. Some of the computer based assessment packages have the ability to calculate the facility value and discrimination index for each question in the test. The teacher can use the statistical analyses of each question to identify which part of the course has a problem or weakness that needs some improvement. Moreover, from the statistical analyses, the teacher may discover that his/her teaching method needs to be changed or modified to improve the quality of the learning process.

g. Question banks. One possibility for electronic generation of tests, is the selection of questions from a bank. Teachers can produce a bank of questions and select from the bank a test for revision and self - assessment purposes. The bank can also be used to meet a test specification.

h. Rapid recording and analysing of the test data. It is the most immediately obvious and most easily accessible use of technology to assist assessment. Some computer based assessment packages have the ability to store the results that can be transferred from the package to different statistical packages and vice versa.

2.3.3 Limitations of Computer Based Assessment:

Computer based assessment has several limitations that prevent teachers from using it as a tool for assessment (Leonard *et al.*, 1997).

a. Computer literacy. Not all the students are familiar with a computer. Although most of the computer based assessment packages are easy to use, it is difficult to make sure that all the students have the skills to use the computer especially if it is used for summative assessment. An introductory course in information technology at the beginning of the course may be necessary to eliminate or reduce this limitation.

b. Reliability of the system. A fault in the computer can cause big problems for students and may result in the loss of important data. It is important that somebody should be able to react quickly if there is any fault in the computer system. This problem can be worse if the computer machines are used to deliver the test for summative purpose.

c. Concentration problems and headaches. This problem is related to the computer user who has health problems when using a computer. Some students have health problems in spending quite a lot of time in front of a computer, such as headaches, allergies........etc. Therefore, teacher should make sure that, if he is going to use the package for summative purpose, no student has any health problem with computers. Any assessment process must not physically harm any student.

d. Security considerations. This point is related to the security of using computers as a tool of delivering assessment to students. There are possibilities that:

- The student which is supposed to sit the test, is the one in the front of the computer.
- Students copying from each other's visual display screens.
- Students may gain entry to files where the marks are stored.

e. Availability of the computers for the whole class. In most universities, the number of computers is not enough to deliver the test to the whole class at the same time. This is really a big problem in using computers for summative assessment. For example, 600 students require 600 computer machines to do the test at the same time.

2.3.4 Examples of Computer Based Assessment Software Packages:

2.3.4.1 Question Mark (QM):

i. Introduction:

Question Mark (QM) for Windows is one of the world's leading computer software packages for computerising tests, assessments and surveys under Windows. It consists of four components (Dempster, 1994):

1. Designer for creating questions.

2. Presenter for answering the questions.

3. Snapshotter for capturing screen images from other Windows applications.

4. Reporter for data analysis and report marking of answer scores. Reporter can produce the test reports in three forms:

- List: test report includes the name of the test, user name, score and data for each answer file selected.
- Summary: test report includes the maximum, minimum and average score and time of completion of selected answer files.
- Full: test report includes the details of the answers for each question.

In terms of the statistical factors, the package has the ability to calculate the average score, standard deviation, facility value and the discrimination index for each question of the test.

ii. Types of Questions:

Question Mark (QM) enables teachers to ask different types of questions in any topic in any subject. Examples of these questions are multiple choice questions, multiple responses, matching questions, true/false and numeric questions (Adams *et al.*, 1998). Nonetheless, most of these questions are MCQs or a modification of MCQs. They have all the disadvantages of the multiple choice questions mentioned earlier.

During the process of setting the test, the test designer should define the way the questions are presented, which answer is right and the type and time the feedback appears to students in the computer screen.

iii. How Does Question Mark (QM) Work?

In order for the examiner to design the test, he must use the Question Mark designer. Students cannot see the set up of the test. They can use the test only from the presenter. Introductory information, at the beginning of the test, is important to give students the full information about the test and what is required from them to answer the questions. The package allows the students to use the test several times unless the test designer has set up the test for only one time usage. Furthermore, the package allows students to move forward and backward in the test or to jump within the test allowing different directions to be taken depending upon the answers chosen.

The test designer determines the number of points allocated to each question. The score may be positive, zero or negative. Both the scores and the feedback can be displayed after each question or at the end of the test if the package is used for self - assessment or formative purposes. If the test is for self - assessment or formative purposes, then the immediate feedback will be very valuable. The time allocated to finish the test can also be determined by the test designer. He may allow students to complete the test when the student is ready or to specify a time by which each student must finish the test.

One advantage of Question Mark (QM) for Windows is that it allows for randomization. It means that the test designer could present different versions of the same question to different students at the same time. This can be done very easily by setting up the randomize facility during the process of designing the test. However, Question Mark does have some limitations like, for example, its limitation in terms of the types of questions (Leonard *et al.*, 1999).

2.3.4.2 Tripartite Interactive Assessment Delivery System (TRIADS):

i. Introduction:

This is a new computer based assessment software package. It has been developed in operation since 1992 at the University of Derby in collaboration with some departments at the University of Liverpool and the Open University. It is still under development and more types of questions will be added. This package seems to be more powerful compared to the Question Mark (QM). It also, like Question Mark (QM) can be run under Windows.

ii. Aim of the TRIADS Project:

The TRIADS project aims to improve the quality of learning by (TRIADS Web Site, 1999).

- Ensuring objective and reproducible assessment of learning outcomes.
- Providing immediate feedback to students.
- Saving time in marking and mark processing.
- Freeing up time for teachers to assess the quality of their courses by quantitative analysis of learning outcomes.

iii. Types of Questions:

The TRIADS system provides the facility to develop a wide range of question styles (Mackenzie, 1999). Examples of these questions are as follows:

1. Move object labelling and building.
2. Multiple choice/response.
3. Move object sequence.
4. Move object on scale.
5. Advance multiple choice/response.
6. Text - numeric entry.
7. Graph plotting.
8. Hot - spot.
9. Field - simulation + hot - spot + draw object + multiple choice questions.
10. Curve line drawing.

iv. The Scoring:

Scoring in TRIADS is flexible and scores for individual selections are teacher - defined. Questions may be weighted and negative scores may be optionally carried - over if desired by the teacher.

v. The Future of Future:

TRIADS package is expected to be used by many institutions in the coming years as a powerful tool for formative, self - assessment and summative assessment. For example, the Institute of Biomedical and Life Sciences (IBLS) of the University of Glasgow tried the system in one of the courses (Genetics Module) by setting up different types of question such as MCQs and sequencing (Sutcliffe *et al.*, 2000).

2.4 Conclusions of Chapter Two:

It can be concluded from the discussions in this chapter, that each assessment method has advantages and disadvantages (Table 2.5). Although there is a variation in the advantages and disadvantages of each method, there should not be a reliance on one assessment method to make a judgement on the pupils' performance. Billington (1997) concluded in his study that diversity of assessment strategies is needed to assess pupils fairly and avoid discrimination. Kellington *et al.* (1980) point out that achievement of many kinds of objectives cannot be investigated by a single type of question. A variety of techniques is required and each should be appropriate for the ability range of pupils.

Furthermore, if any assessment method is used too frequently, pupils can learn to respond automatically. For example, the one who is familiar with MCQs will get an inflated score because of skills, gained through experience, of ruling out distractors, willingness to guess, or better management of time. The problem here is not having the skills but the questions are not testing the understanding of what the test is intended to measure.

Thus, using diverse methods of assessment will improve the quality of learning in two ways (White *et al.*, 1992).

- They promote greater awareness by students of the quality of their learning.
- Shows students that subjects involve a range of learning outcomes and are far more interesting in consequence.

More research and development of computer software for computer based assessment are needed to find good strategies of formative and summative purposes of assessment and go beyond using only fixed response questions. Black (1993) recommends that with further development software designed to serve good formative strategies could be a powerful aid. The introducing of such technology will affect any school or institution programme. However, such introduction should not sacrifice the quality of any assessment process.

Table 2.5: Comparison between some fixed and open response methods of assessment

	Fixed Response					Open Response
	MCQs	**T/F and Linked T/F**	**Matching**	**Venn Diagrams**	**SCGs**	**Essay**
Outcomes Measured	Good for measuring low levels of understanding.	Good for measuring low levels of understanding.	Good for identifying the relationships between two things.	Good for measuring the categorisation.	Good for measuring different levels of understanding.	Good for measuring high levels of understanding.
Sampling of Content	A wide selection of topics can be tested in a short period.	A wide selection of topics can be tested in a short period for normal T/F. But not the case for Linked T/F.	Cannot test a large sample of the course content.	Cannot test a large sample of the course content.	Can test a large sample of the course content.	Cannot test a large sample of the course content.
Testing Large Number of Students	Good method for testing large number of students.	Can be used to test large number of students.	Can be used to test large number of students.	Can be used to test large number of students.	Can be used to test large number of students.	Not suitable to test large number of students.
Preparation of Items	Preparation of good items is difficult and time consuming.	They are relatively easy to write and construct. However, it is difficult to write Linked T/F.	Difficult to prepare questions that can be used to test the higher levels of cognitive domain.	Difficult to prepare because it is only suitable for some parts of the curriculum.	They are much easier than MCQs to write and construct.	Easy to set and prepare.

Table 2.5: Comparison between fixed and open response methods of assessment (Continued)

	Fixed Response					Open Response
	MCQs	**T/F and Linked T/F**	**Matching**	**Venn Diagrams**	**SCGs**	**Essay**
Validity	There is a problem of the validity.	There is a problem of the validity. However, the validity of Linked T/F seems to be better than normal T/F.			Seems to be better in the validity than MCQs.	They may be valid instruments than MCQs.
Reliability	Have a satisfactory test inter - marker reliability.	Have a satisfactory test inter - marker reliability.	Have a satisfactory test inter - marker reliability.	Have a satisfactory test inter - marker reliability.	Have a satisfactory test inter - marker reliability.	They are a less reliable compared to the objective questions.
Scoring	Can be scored easily, objectivity, and quickly and lend themselves to machine scoring.	Can be scored easily and objectivity. A computer programme can be designed to score Linked T/F.	Can be scored rapidly, accurately and objectively.	Can be scored easily and objectively.	Objectively markable. A computer programme can be designed to score them.	It takes long time to score. There is a problem of subjectivity in scoring.

Table 2.5: Comparison between some fixed and open response methods of assessment (Continued)

	Fixed Response					Open Response
	MCQs	**T/F and Linked T/F**	**Matching**	**Venn Diagrams**	**SCGs**	**Essay**
Guessing	There is a problem of guessing and cueing effects.	The normal T/F can be answered correctly by blind guessing.	Relatively low chance of guessing if the number of responses is more than the number of premises.	There is low chance for guessing.	The indication so far shows that there is low chance for guessing.	There is no chance for guessing. The guessing comes only in a question spotting.
Credit for Partial Knowledge	Usually there is no credit for partial knowledge.	No credit for partial knowledge in the normal T/F. However, in the Linked T/F student can gain credit for his partial knowledge.	There is a credit for each correct matching between the premise and its response.	There is a credit for partial knowledge.	There is a credit for partial knowledge.	There is a credit for partial knowledge.
Study Habit	Encourage students to adopt surface approaches to studying.	Encourage students to study oversimplified statements of truth and factual details.	Encourage students to memorize isolated facts.		Encourage students to adopt deep approaches to studying.	Encourage students to adopt deep approaches to studying.

CHAPTER THREE: FIELD RESEARCH AND RESULTS (1) - PART (A)

CHAPTER THREE: FIELD RESEARCH AND RESULTS (1) – PART (A)

3.1 Introduction:

This chapter is going to describe the methods which were used to tackle the first part of the research problems and to present the findings of this research. It consists of two parts; **A** and **B**. We will start with part (**A**) which is attempting to discuss issues related to multiple choice questions.

At the beginning, it is important that the term 'facility value' of a question is defined and understood. Facility value (F.V) can be defined as the percentage of pupils in a given sample answering the item correctly (Nitko, 1983; Friel *et al.*, 1988). In other words, it is the proportion of pupils answering the item correctly from the total number of pupils answering the item or taking the test. In the case of open response questions or other questions, which are scored out of several marks, facility value is often the mean mark for the question converted to a percentage of the maximum mark (Hall, 1988).

An 'acceptable' facility value depends upon the purpose of the test. In terms of norm - referenced assessment, some educators such as Friel *et al.* (1988) state that the facility value of the question should lie in the range of 35 - 85 percent. If the value is below 35% it means that the question is so difficult that most of the students could not answer it correctly. On the other hand, if the value is above 85% it means that the question is so easy that even the weak students answered it correctly. This range of facility value will tend to produce a paper, which discriminates well.

However, there are occasions where one would expect all pupils to get the correct answer. This situation is applied for criterion - referenced assessment where the facility value of the question, ideally, is equal to 1.

3.2 Fixed Response: Multiple Choice Questions

3.2.1 Changing the Position of the Most Plausible Distractor (Strong Distractor):

3.2.1.1 The Aim of the Research:

The aim of this part of the research is to investigate the effect on the facility value of changing the position of the most plausible distractor (strong distractor) of the question when it is placed next to, or away, or further away from the key answer. The most plausible distractor, simply, is the distractor, which is selected most among other distractors in some pre – test.

3.2.1.2 The Research Hypothesis:

This part of the research aims to test the following hypothesis:

Hypothesis 3.1: There is no significant difference in the item's facility value when the most plausible distractor is placed next to, or away, or further away from the key answer.

3.2.1.3 The Method of Research:

In the first two years of the research period, two multiple choice tests were applied to first year university students in biology. These were based on Evidence for Evolution and Transport - Gas Exchange. The former was applied in the third term of the 1997/98 academic year, whereas the latter was applied in the second term of 1998/99 academic year. These two tests were administered to students for self – assessment purpose which means that they are voluntary.

i. Test One: Evidence for Evolution:

a. Description of the Test:

The Evidence for Evolution test consists of thirty (30) multiple choice questions (Appendix 1.1). The original version of the test was used plus two other versions which were designed for the purpose of the research. The researcher used the results of the last academic year (1996/97) of the test to determine whether there is a distractor that could appear stronger in the students' responses. Then, the researcher designed two more versions of the original test based upon these results. The only difference between the two versions comparing to the original one is that the most plausible distractor (strong distractor) was away or further away from the key answer. Questions twelve and seventeen of the test were not exposed to any change and were used as internal markers to check if the student samples were comparable.

Example from Evidence for Evolution Test:

The Stem*: Anatomic structures that show similar function but dissimilar embryonic and evolutionary background are said to be..............*

(Original Version) *Strong Distractor Next To the Key Answer*	(Version One) *Strong Distractor Away from the Key Answer*	(Version Two) *Strong Distractor Further Away from the Key Answer*
a. primitive.	a. primitive.	a. homologous. *
b. polyphyletic.	b. homologous. *	b. polyphyletic.
c. homologous. *	c. polyphyletic.	c. primitive.
d. analogous. •	d. analogous. •	d. analogous. •
e. monophyletic.	e. monophyletic.	e. monophyletic.

Note:-

• = the key answer.

** = the most plausible distractor (strong distractor).*

b. The Sample:

The sample was the first year university students of the University of Glasgow who were taking first year biology modules. All the test versions were administered to students using the computer based assessment package **'Question Mark'** (To know more about this package, see Chapter Two - section 2.3.4.1). The number of students who answered each question in the three versions varied from one question to another but the minimum number of students who answered a question was 34 (Table 3.1).

The computer package randomly selected the version of the question which was presented to a particular student. Therefore, any one student could answer the questions which belong to all three test versions, thus further randomizing the sample.

Table 3.1: Number of students who answered the three versions of the Evidence for Evolution test

	Number of Students				Number of Students		
Item Number	**Original Version**	**Version One**	**Version Two**	**Item Number**	**Original Version**	**Version One**	**Version Two**
1	55	71	69	**16**	44	46	55
2	65	64	51	**17**	39	53	52
3	67	50	59	**18**	43	52	47
4	54	52	66	**19**	38	47	57
5	58	56	58	**20**	43	52	44
6	53	50	67	**21**	40	55	43
7	50	65	54	**22**	41	52	44
8	63	52	50	**23**	44	34	59
9	52	63	46	**24**	42	50	45
10	60	45	52	**25**	43	50	42
11	46	61	45	**26**	45	46	44
12	43	52	52	**27**	51	42	41
13	47	57	42	**28**	53	44	47
14	52	50	42	**29**	34	50	50
15	46	51	47	**30**	53	43	37

ii. Test Two: Transport - Gas Exchange:

a. Description of the Test:

The aim of this test is give further evidence about the effect of changing the position of the most plausible distractor on the facility value of the MCQs when it is placed next to, or away, or further away from the key answer. The researcher used the results of the last academic year (1997/98) of the test to find out the most plausible distractor. Based upon the results obtained from the last academic year, three versions were designed.

- The original version where the strong distractor was placed next to the key answer.
- Version one where the strong distractor was placed away from the key answer.
- Version two where the strong distractor was placed further away from the key answer.

The original version and version one consist of twenty one (21) multiple choice questions (Appendix 1.2) whereas version two consists of nineteen (19) questions. Questions fifteen (15) and eighteen (18) were not found in version two due to the fact that these two questions had only three options. The number of options in each item varied, some having three or four or five. As in the previous test, the only difference between the three versions was the position of the most plausible distractor (strong distractor) to the key answer. No change was made to questions four and twelve of the test to act as internal markers.

Example from Transport - Gas Exchange Test:

The Stem: *In which animal does blood flow through vessels from the respiratory organ back to the heart before circulating through the rest of the body?*

(Original Version) *Strong Distractor Next To the Key Answer*	(Version One) *Strong Distractor Away from the Key Answer*	(Version Two) *Strong Distractor Further Away from the Key Answer*
a. frog. •	a. frog. •	a. frog. •
b. fish. *	b. annelid.	b. mollusc.
c. annelid.	c. fish. *	c. annelid.
d. mollusc.	d. mollusc.	d. fish. *
e. insect.	e. insect.	e. insect.

Note:-

• = *the key answer.*

* = *the most plausible distractor (strong distractor).*

b. The Sample:

In this test, the computer package for self - assessment (**Question Mark**) was also used to administer the test. The computer package randomly altered version of the question which was presented to the student in the computer screen. The number of students who answered each question in the three versions of this test varied from one question to another but it was not fewer than eighty eight (88) students (Table 3.2).

Table 3.2: Number of students who answered the three versions of Transport - Gas Exchange test

	Number of students		
Item Number	**Original Version**	**Version One**	**Version Two**
1	131	154	156
2	159	135	143
3	136	161	135
4	142	142	140
5	143	144	133
6	156	140	122
7	136	136	134
8	125	128	147
9	128	149	121
10	116	135	118
11	120	124	115
12	116	103	134
13	115	118	118
14	133	105	106
15	173	169	-
16	100	133	108
17	108	113	110
18	167	166	-
19	114	104	107
20	92	114	105
21	107	88	113

3.2.2 The Results of Changing the Position of the Most Plausible Distractor (Strong Distractor):

3.2.2.1 Introduction:

The results of some previous studies on the effect of changing the position of the most plausible distractor in the facility value of the MCQs (Marcus, 1963; Friel, 1979) showed that changing the position of the most plausible distractor brings a significant effect in the facility value of the MCQs. However, we would like to see if our results agree with (support) or disagree with Marcus (1963) and Friel (1979).

3.2.2.2 Test One: Evidence for Evolution:

The facility value of each item of the three versions was calculated and tabulated in Table 3.3.

Table 3.3: The facility value of each item of the three versions of the Evidence for Evolution test

	Facility Value				Facility Value		
Item Number	**Next To**	**Away**	**Further Away**	**Item Number**	**Next To**	**Away**	**Further Away**
1	0.44	0.61	0.58	**16**	0.57	0.67	0.55
2	0.63	0.80	0.78	**17**	0.96	0.98	0.97
3	0.85	0.94	0.90	**18**	0.65	0.73	0.68
4	0.44	0.46	0.44	**19**	0.71	0.66	0.68
5	0.62	0.57	0.59	**20**	0.74	0.52	0.57
6	0.83	0.86	0.79	**21**	0.93	0.89	0.88
7	0.58	0.58	0.59	**22**	0.78	0.77	0.80
8	0.56	0.62	0.46	**23**	0.73	0.71	0.69
9	0.44	0.32	0.28	**24**	0.81	0.78	0.80
10	0.27	0.33	0.31	**25**	0.60	0.64	0.55
11	0.54	0.46	0.49	**26**	0.36	0.30	0.45
12	0.58	0.56	0.57	**27**	0.51	0.50	0.56
13	0.81	0.75	0.93	**28**	0.67	0.61	0.55
14	0.79	0.70	0.79	**29**	0.76	0.84	0.86
15	0.59	0.59	0.47	**30**	0.49	0.51	0.49

From Table 3.3 we can see that changing the position of the most plausible distractor (strong distractor) affects the facility value of the MCQs. It supports the results of Marcus (1963) and Friel (1979) studies. The following are examples.

Item one of the test has a range of 17% in the facility value between the original version where the strong distractor was placed next to the key answer and version one where the strong distractor was placed away from the key answer (F.V of original version = 0.44 and of version one = 0.61). This value decreased to 14% between the original version and version two when the strong distractor was located further away from the key answer (F.V of version two = 0.58).

Item two also suffered from changing the position of the most plausible distractor (strong distractor). The difference in the facility value between the original version and version one is 17%. However, there is not much difference in the facility value when the strong distractor was placed further away from the key answer in version two (between version one and version two).

Item nine is another example that has a difference in the facility value from one version to another by 10% or more. The difference in the facility value between the original version and version one is by 12% (F.V of the original version = 0.44 and of version one = 0.32). On the other hand, the difference in the facility value between the original version and version two is by 16% (F.V of the original version = 0.44 and of version two = 0.28).

Item twenty has a big difference in the facility value among all the items. The difference in the facility value between the original version where the strong distractor was placed next to the key answer and version one where the strong distractor was placed away from the key answer is by 22% (F.V of the original version = 0.74 and of version one = 0.52). Nonetheless, when the most plausible distractor was placed further away from the key answer in version two the facility value decreased by 5% to become 17% (F.V of the original version = 0.74 and of version two = 0.57). The difference in the facility value between one version to another by 10% or more also happen in items **8, 13, 15, 16, 26** and **29** (see Table 3.3).

Furthermore, a closer look to the Table 3.3 the following can be concluded:

1. If we compare the facility value of the items across the three versions of the test, we can see almost a third of the items have higher facility value when the most plausible (strong distractor) was placed away from the key answer (version one). The following is the statistics of this observation:

- In version one, twelve items out of thirty (40%) have higher facility value than the original version and version two.
- In the original version, ten items out of thirty (33%) have higher facility value than version one and version two.
- In version two, six items out of thirty (20%) have higher facility value than the original version and version one.
- The remaining two items (0.7%) have high facility values in more than one version.

2. There is no clear pattern of the difference in the magnitude of the facility value when the most plausible distractor is placed in different positions to the key answer (i.e. higher or lower compared to the facility value of the original version) across the test versions. However, the best pattern the researcher could identify is at the cut off point 0.61. If the facility value (F.V) of the original version (*most plausible distractor is located next to the key answer*) is larger than or equal to 0.61 (F.V $\geq$ 0.61), then the facility value of version one (*most plausible distractor is located away from the key answer*) and version two (*most plausible distractor is located further away from the key answer*) is overall lower although this is not clearly shown in version two. On the other hand, if the facility value of the original version is smaller than 0.61 (F.V < 0.61), then the facility value of version one (away) is overall higher than the facility value of the original version but the facility value of version two (further away) is overall lower than the facility value of the original version. It must be that there are other factors which affect the difference in the facility value behind the position of the most plausible distractor. By checking the questions of the test in depth, the results showed that the facility value of the question is affected by unequal (uneven) length of the options. If all options of the item are of equal length, then it was found that the facility value would be higher when the most plausible distractor (strong distractor) is located next to the key answer. Oppositely, if the options are not of equal length, then the longer or the shorter option sometimes draws the responses from the key answer.

3. Most of the differences in the facility value occur between the original version and version two. Nine items out of eighteen (50%) of those items which have the difference in the facility value by 10% or more were found between the original version and version two. Five items out of eighteen (28%) were found between the original version and version one. The remaining four items (22%) of those items which have the difference in the facility value by 10% or more were found between version one and version two.

Are the differences in the facility value of those items, which have the difference in the facility value by 10% or more, statistically significant? The method designed by Kellett (1978) was used to calculate the level of significance (see Appendix 1.3 for Kellett's method of calculating the level of significance). Table 3.4 shows the items which are statistically significant.

Table 3.4: The level of significance of some items of Evidence for Evolution test

	Items		
Level of Significance	**Original Version and Version One**	**Original Version and Version Two**	**Version One and Version Two**
5%	1,2, 16, 20	2, 9, 15, 20, 28	8, 13, 15, 26
1%	2, 20	9, 15, 20	8, 26

From the table we can see that ten items out of seventeen of those items, which have the difference in the facility value by 10% or more, are statistically significant. Therefore, hypothesis **3.1**, which states that '*there is no significant difference in the item's facility value when the most plausible distractor is placed next to, or away, or further away from the key answer*' can be rejected for items **1, 2, 8, 9, 13, 15, 16, 20, 26** and **28.**

3.2.2.3 Test Two: Transport - Gas Exchange:

The aim of this test is give a further evidence about the effect of changing the position of the most plausible distractor on the facility value of the MCQs when it is placed next to, or away, or further away from the key answer. To do so, the facility value of each item of the three test versions is shown in Table 3.5.

Table 3.5: The facility value of each item of the three versions of Transport – Gas Exchange test

	Facility Value		
Item Number	**Next to**	**Away**	**Further Away**
1	0.83	0.81	0.83
2	0.54	0.50	0.50
3	0.74	0.73	0.77
4	0.72	0.71	0.70
5	0.90	0.89	0.89
6	0.58	0.59	0.51
7	0.44	0.41	0.34
8	0.49	0.59	0.46
9	0.78	0.78	0.73
10	0.84	0.80	0.86
11	0.52	0.56	0.55
12	0.61	0.60	0.62
13	0.58	0.55	0.66
14	0.77	0.66	0.67
15	0.90	0.85	-
16	0.32	0.35	0.36
17	0.57	0.54	0.55
18	0.41	0.45	-
19	0.50	0.53	0.45
20	0.35	0.46	0.44
21	0.51	0.41	0.51

The same results, which were obtained for some items in the previous test, were also found in this test. Some items have a difference in the facility value between one version to another by 10% and more.

From Table 3.5, item seven has a difference in the facility value between the original version where the most plausible distractor was located next to the key answer and version two where the most plausible distractor was located further away from the key answer is by 10% (F.V of the original version = 0.44 and version two = 0.34).

Another example is item eight of the test. In this item, the difference in the facility value between the three versions is by 10% and more. The facility value of the original version is 0.49 and 0.59 for version one. The difference between the two versions is 10%. However, when we look at the facility value of version two and compared it with the facility value of version one we can see that the difference between the two versions is 13% (F.V of version one = 0.59 and of version two = 0.46).

The other two examples of those items, which have the differences in the facility value from one version to another by 10% or more, are thirteen and fourteen. In item thirteen, the difference in the facility value between version one where the most plausible distractor was located away from the key answer (F.V = 0.55) and version two where the most plausible distractor was located further away from the key answer (F.V = 0.66) is by 10%. The difference in the facility value between the original version and version one of item fourteen is by 10%. More examples are items twenty and twenty one (see Table 3.5 above).

Moreover, from the same table (Table 3.5) we reach slightly different results from what we found in the Evidence for Evolution test in terms of which test version has more items with higher facility value. The results showed that seven out of twenty one have higher facility value when the most plausible (strong distractor) was placed next to the key answer (original version). On the other hand, six items out of twenty one was found higher facility value when the most plausible distractor was placed away from the key answer (version one). Finally, it was found that five items out of twenty one have higher facility value when the most plausible distractor was placed further away from the key answer (version two).

Also, there is no clear pattern of the difference in the magnitude of the facility value when the most plausible distractor is located in different positions to the key answer (i.e. higher or lower compared to the facility value of the original version) across the test versions. Using the same range which was used for Evidence for Evolution test, it was found that when the facility value (F.V) of the original version (*most plausible distractor is located next to the key answer*) is larger than or equal to 0.61 (F.V $\geq$ 0.61), then the facility value of version one (*most plausible distractor is located away from the key answer*) and version two (*most plausible distractor is located further away from the key answer*) is overall lower. However, when the facility value of the original version is smaller than 0.61 (F.V < 0.61), then the facility value of version one (away) is overall higher than the facility value of the original version but the facility value of version two (further away) is overall lower than the facility value of the original version. Furthermore, the results showed that if the options of the item are all in equal length, then the facility value is higher when the most plausible distractor (strong distractor) is located next to the key answer. On the other hand, if the options are not of equal length then it seemed to be that the longer or the shorter option plays a part in affecting the facility value of the item.

In this test, it was found that most of the difference in the facility value occurred between the original version and version one and between version one and version two. Half of the items (four out of eight) of those items, which have the difference in the facility value between one version to another by 10% or more, were between the original version and version one, three items out of eight (38%) between version one and version two and only one item between the original version and version two.

We would like to know now which of the above items have a statistically significant difference in the facility value. Again Kellett's (1978) method of calculating the level of significance was applied. Table 3.6 shows the items which are statistically significant.

Table 3.6: The level of significance of some items of Transport - Gas Exchange test

	Items		
Level of Significance	**Original Version and Version One**	**Original Version and Version Two**	**Version One and Version Two**
5%	8, 20	7, 14, 20	8, 13
1%	8	7	8, 13

From the table above, few items were found statistically significant at 1% or 5% levels. Accordingly, hypothesis **3.1** can be rejected for items **7, 8, 13, 14** and **20**.

3.2.3 Changing the Position of the Key Answer:

3.2.3.1 The Aim of the Research:

The aim of this part of the research is to examine whether altering the position of the key answer affects the facility value of the multiple choice questions.

3.2.3.2 The Research Hypothesis:

The following hypothesis was tested in this part of the research:

Hypothesis 3.2: There is no significant difference in the item's facility value when the position of the key answer has been altered. However, this automatically alters the proximity of the most plausible distractor (strong distractor).

3.2.3.3 The Method of Research:

In order to examine the effect of changing the position of the key answer on the facility value of the MCQs, two multiple choice tests were applied to the first year university students in biology. These were Introduction to Animal Physiology and Origin of Species. Introduction to Animal Physiology test was applied in the third term of the 1997/98 academic year, whereas Origin of Species test was applied in the third term of 1998/99 academic year. Again, the two tests were used for self – assessment purpose which means that they are voluntary.

i. Test One: Introduction to Animal Physiology:

a. Description of the Test:

This test consists of eighteen (18) multiple choice questions (Appendix 2.1). The researcher used the original version and produced another four versions. The wording of the five versions was identical except for the position of the key answer. Item number ten of the test was kept without change for all test versions to be used as an internal marker. The test was administered to students using the computer based assessment package **'Question Mark'**. This package, as was mentioned above, allows for randomization. Consequently, it is unlikely that any one student will answer the items of only one version.

Example from Introduction to Animal Physiology Test:

The Stem: *Which type of muscle is responsible for peristalsis along the digestive tract?*

(Ver. Three) *Key Answer in (a) Position*	**(Ori. Version)** *Key Answer in (b) Position*	**(Ver. Two)** *Key Answer in (c) Position*	**(Ver. One)** *Key Answer in (d) Position*	**(Ver. Four)** *Key Answer in (e) Position*
a. smooth. •	a. cardiac.	a. cardiac.	a. cardiac.	a. cardiac.
b. voluntary.	b. smooth. •	b. voluntary.	b. voluntary.	b. voluntary.
c. cardiac.	c. voluntary.	c. smooth. •	c. striated.	c. skeletal.
d. striated.	d. striated.	d. striated.	d. smooth. •	d. striated.
e. skeletal.	e. skeletal.	e. skeletal.	e. skeletal.	e. smooth. •

Note:-

• *the key answer.*

b. The Sample:

The test was applied to the first year university students of the University of Glasgow, who were taking biology modules. The number of students who answered each item varied but was not less than 68 (Table 3.7).

Table 3.7: Number of students who answered the different versions of the Introduction to Animal Physiology test

	Number of students				
Item Number	**Original Version**	**Version One**	**Version Two**	**Version Three**	**Version Four**
1	132	151	138	165	171
2	171	204	169	204	
3	174	180	199	177	
4	124	142	156	154	149
5	152	148	136	146	136
6	186	170	183	175	
7	144	153	121	142	142
8	162	138	146	117	134
9	72	78	72	72	68
10	137	130	136	141	141
11	153	175	193	162	
12	129	141	138	141	128
13	140	185	172	176	
14	146	122	140	128	136
15	126	136	127	132	151
16	137	123	124	137	151
17	139	124	123	140	130
18	122	139	121	137	127

ii. Test Two: Origin of Species:

a. Description of the Test:

The purpose of this test is give further evidence on the effect of changing the position of the key answer on the facility value of the MCQs. It consists of twenty one (21) multiple choice test (Appendix 2.2). The researcher used the original version and produced another four versions. Again the only difference between the five versions was the position of the key answer. Item six of the test was kept the same to be used as an internal marker. The same computer package 'i.e. **Question Mark**' was used to administer the test.

Example from the Origin of Species Test:

The Stem: *The only taxonomic category that actually exists as a discrete unit in nature is the..................*

(Ver. Two) *Key Answer in (a) Position*	(Ori. Version) *Key Answer in (b) Position*	(Ver. One) *Key Answer in (c) Position*	(Ver. Three) *Key Answer in (d) Position*	(Ver. Four) *Key Answer in (e) Position*
a. species. •	a. class.	a. class.	a. class.	a. class.
b. class.	b. species. •	b. family.	b. genus.	b. phylum.
c. family.	c. family.	c. species. •	c. family.	c. family.
d. genus.	d. genus.	d. genus.	d. species. •	d. genus.
e. phylum.	e. phylum.	e. phylum.	e. phylum.	e. species. •

Note:-

• the key answer.

b. The Sample:

The sample was composed of first year biology students of the University of Glasgow. The number of students who answered the test versions was different from one version to another (Table 3.8).

Table 3.8: Number of students who answered the different versions of the Origin of Species test

	Number of Students				
Item Number	Original Version	Version One	Version Two	Version Three	Version Four
1	137	112	108	100	119
2	115	107	111	117	117
3	106	102	130	109	117
4	101	116	101	111	120
5	108	128	99	100	109
6	107	92	121	98	115
7	103	101	107	103	115
8	118	107	103	95	100
9	99	105	114	100	100
10	104	103	105	100	104
11	101	109	101	116	88
12	91	107	106	111	97
13	116	96	101	108	97
14	103	89	106	114	95
15	107	97	98	95	106
16	99	103	86	104	108
17	116	92	99	110	81
18	95	88	118	93	99
19	103	97	99	100	87
20	97	104	83	91	108
21	91	93	109	87	100

3.2.4 The Results of Changing the Position of the Key Answer:

3.2.4.1 Introduction:

The results of some previous studies on the effect of changing the position of the key answer (McNamara *et al.*, 1945; Payne, 1951; Bluch, 1984; Ace *et al.*, 1973; Cizek, 1994) reported that the facility value of the MCQs also suffers from changing the position of the key answer. The current research is trying to find out if the similar result will be obtained.

3.2.4.2 Test One: Introduction to Animal Physiology:

The facility value of each item of the five versions of the test was calculated and presented in Table 3.9 below.

Table 3.9: The facility value of each item in different positions of the key answer of Introduction to Animal Physiology test

	Facility Value				
Item Number	**In (a) Pos.**	**In (b) Pos.**	**In (c) Pos.**	**In (d) Pos.**	**In (e) Pos.**
1	0.83	0.80	0.75	0.78	0.79
2	0.61	0.60	0.56	0.57	
3	0.86	0.83	0.84	0.83	
4	0.62	0.68	0.70	0.67	0.71
5	0.60	0.64	0.66	0.65	0.62
6	0.79	0.67	0.63	0.61	
7	0.62	0.69	0.73	0.68	0.59
8	0.81	0.81	0.86	0.82	0.82
9	0.65	0.46	0.60	0.58	0.51
10	0.56	0.57	0.57	0.57	0.56
11	0.52	0.54	0.67	0.55	
12	0.77	0.71	0.67	0.67	0.72
13	0.77	0.72	0.69	0.74	
14	0.91	0.89	0.91	0.92	0.84
15	0.91	0.87	0.94	0.88	0.91
16	0.64	0.66	0.64	0.73	0.58
17	0.66	0.64	0.59	0.63	0.69
18	0.90	0.91	0.86	0.91	0.88

It is obvious from Table 3.9 that some items have a difference in their facility value from one version to another by 10% or more. Examples are items **6,7, 9,11,12** and **16.**

The facility value of item six is 0.79 when the key answer was placed in a position. Then, it started to decrease steadily to become 0.67 when the key answer was placed in position **b**, 0.63 when the key answer was placed in position **c** and 0.61 when the key answer was placed in position **d**.

Item seven is another example which has a fluctuation in its facility value from one version to another. The facility value increased from 0.62 when the key answer was placed in position **a** to 0.69 when the key answer was placed in position **b** to 0.73 when the key answer was placed in position **c**. Then, it started to decrease to 0.68 when the key answer was placed in position **d** and to 0.59 when the key answer was placed in **e** position.

Another example is item nine of the test. In this item, the facility value is 0.65 when the key answer was written in position **a**. However, when the key answer was written in position **b** the facility value decreased by 19% (F.V in **a** position = 0.65 and in **b** position = 0.46). In the original version, where the key answer was written in position **c**, the facility value started to increase and became 0.60. There is not much difference when the key answer was written in position **d** compared when it was written in **c** position.

This kind of fluctuation can also be found in item sixteen. In this item, the facility value is 0.64 when the key answered was placed in position **a**. Then, it started to increase to 0.66 when the key answer was placed in position **b**. On the other hand, when the key answer was placed in position **c**, the facility value started to decrease and became 0.64. The facility value increased to 0.73 when the key answer was placed in position **d**. Finally, when the key answer was placed in position **e**, the facility value decreased to 0.58.

Another conclusion which we can draw from Table 3.9 is that the facility value is higher, in most items, when the key answer was placed in **a** or **c** positions. In other words, it seemed to be that students found the questions easier when the key answer was placed in position **a** or in position **c**. It was found that seven out of eighteen items (39%) have highest facility value when the key answer was placed in position **a**. Five out of eighteen items (28%) were found that their facility value is highest when the key answer was placed in position **c**.

However, are the items that have a difference in the facility value of 10% or more, statistically significant? To find out an answer of this question, Kellett's (1978) method of testing the level of significance was used. The items which are statistically significant are presented in Table 3.10 below.

Table 3.10: The level of significance of some items of Introduction to Animal Physiology test

	Level of Significance	
Between	**5%**	**1%**
Original Version and Version One	7, 17	7, 17
Original Version and Version Two	6, 7, 11	6, 7, 11
Original Version and Version Four	9, 12	9, 12
Version One and Version Two	6, 11, 16	6, 11, 16
Version One and Version Three	9	9
Version Two and Version Three	6, 11	6, 11
Version Two and Version Four	7, 12	12
Version Three and Version Four	9	9

From the table above, we can see that many items are statistically significant at 1% or 5%. Consequently, the null hypothesis (hypothesis **3.2**), which proposes that *'there is no significant difference in the item's facility value when the position of the key answer has been altered'*, can be rejected for items **6, 7, 9, 11, 12, 16,** and **17**.

<u>3.2.4.3 Test Two: Origin of Species:</u>

The purpose of this test, as was mentioned in section 3.2.3.3 of the current chapter, is to find out more evidence on the effect of changing the position of the key answer in the facility value of the MCQs.

The facility value of each item of the five versions of the test was calculated and presented in Table 3.11.

Table 3.11: The facility value of each item in different positions of the key answer of the Origin of Species test

	Facility Value				
Item Number	**In (a) Pos.**	**In (b) Pos.**	**In (c) Pos.**	**In (d) Pos.**	**In (e) Pos.**
1	0.77	0.66	0.69	0.74	0.71
2	0.51	0.50	0.59	0.43	0.46
3	0.62	0.60	0.59	0.62	0.50
4	0.75	0.77	0.79	0.73	0.75
5	0.65	0.69	0.77	0.63	0.62
6	0.53	0.51	0.53	0.52	0.51
7	0.57	0.57	0.60	0.69	0.63
8	0.62	0.62	0.53	0.57	0.49
9	0.69	0.76	0.78	0.68	0.61
10	0.85	0.81	0.73	0.62	0.66
11	0.80	0.84	0.77	0.75	0.73
12	0.83	0.77	0.82	0.79	0.86
13	0.49	0.41	0.45	0.52	0.52
14	0.66	0.69	0.72	0.63	0.64
15	0.72	0.74	0.69	0.67	0.68
16	0.73	0.76	0.76	0.78	0.79
17	0.47	0.57	0.48	0.49	0.57
18	0.22	0.24	0.30	0.29	0.23
19	0.54	0.53	0.61	0.49	0.60
20	0.69	0.75	0.69	0.62	0.75
21	0.56	0.72	0.52	0.48	0.49

As it can be seen from Table 3.11 that changing the position of the key answer has an effect in the facility value of some items of this test. This supports the results which were found from Introduction to Animal Physiology test.

For example the facility value of item two is 0.59 when the key answered was placed in position **c** but it is 0.43 when the key answered was placed in position **d**. The difference in the facility value between the two positions is by 16%.

In item eight, the difference in the facility value between one version to another is by 10%. When the key answer was placed in positions **a** or **b**, the facility value is 0.62. However, when the key answer was moved to position **c**, the facility value decreased to 0.49.

Item ten has a very big difference in the facility value between one version and another. The facility value of the item is 0.85 and 0.81 when the key answer was placed in positions **a** and **b** respectively. But when the key answer was placed in the bottom positions (**d** or **e**), the facility value became 0.62 (in **d** position) and 0.66 (in **e** position). It means that students found item ten easier when the key answer was placed in **a** or **b** positions than when it was placed in **d** or **e** positions.

Another example of this difference is item nineteen of the test. In this item, the facility value is 0.61 when the key answer was placed in **c** position. Then, it decreased to 0.49 when the key answer was placed in position **d** (see Table 3.11 for items **1**, **3**, **5**, **7**, **9**, **13**, **17**, **20** and **21**).

Also in this test and the same as in the Introduction to Animal Physiology test, students found twelve questions out of twenty one easier (higher facility value) when the key answer was placed in **a**, **b** and **c** positions.

After we have seen the items, which have the difference in the facility value by 10% or more, we would like to know which of these items are statistically significant The same method, which was used for Introduction to Animal Physiology test, was applied to test the level of significance of this test. Table 3.12 shows the items which are statistically significant.

Table 3.12: The level of significance of some items of the Origin of Species test

	Level of Significance	
Between	**5%**	**1%**
Original Version and Version One	9, 10, 13, 21	9, 10, 21
Original Version and Version Two	5, 8, 19	5, 19
Original Version and Version Three	9, 13, 20	9, 13, 20
Original Version and Version Four	2	2
Version One and Version Two	5, 10, 19, 21	10, 19, 21
Version One and Version Three	7, 10, 21	7, 10, 21
Version One and Version Four	2, 21	2, 21
Version Two and Version Three	5, 9, 10, 17, 20	5, 9, 10, 20
Version Two and Version Four	8	8
Version Three and Version Four	3, 7	3, 7

The table shows that many of the items, which have the difference in the facility value by 10% and more, are statistically significant. Therefore, hypothesis **3.2** can be rejected for items **2, 3, 5, 7, 8, 9, 10, 13, 17, 19, 20** and **21**.

3.2.5 Examine the "Optical" Explanation:

3.2.5.1 Introduction:

The current research is not only trying to examine if there is a difference in the facility value of the MCQs, when the positions of the key answer and the most plausible distractor have been rearranged, but also to find out an explanation for these differences. We know from the previous sections (sections 3.2.2 and 3.2.4) that the facility value of the MCQs was affected by changing the positions of the key answer and the most plausible distractor. There is no obvious explanation of this problem. It could be due to many factors which are not easily identified. Nevertheless, we can think of one factor which may, to some extent, be the cause of this problem. This factor is something optical. However, it was found also in previous section (section 3.2.2) that the question is affected by unequal (uneven) length of the options. If all options of the item are of equal length, then facility value, in general, would be higher when the most plausible distractor (strong distractor) is located next to the key answer. On the other hand, if the options are not in the equal length, then the longer or the shorter option sometimes draws the responses from the key answer.

In the current research, we are going to investigate the effect of the 'optical' factor on the differences of the facility values of the MCQS. The following paragraphs present how this explanation (factor) was examined practically.

3.2.5.2 The Aims of the Research:

Two aims were addressed in this part of the research:

- Examine if the differences in the facility values of the MCQs, when the positions of the key answer and the most plausible distractor have been changed, is due to an optical effect.
- To encourage pupils/students to discard the weakest distractors, they were given credit for doing so and showing their partial knowledge.

3.2.5.3 The Research Hypotheses:

The following hypotheses were proposed to test in this part of the research:

Hypothesis 3.3: There is no significant difference in the item's facility value when pupils/students are asked to answer the MCQs by two different methods; the normal method (select the correct option) and by the method of removing the two definitely wrong options and then select the correct one from the remaining options.

Hypothesis 3.4: There is no significant difference in the item's facility value, when the position of the key answer is altered, if pupils are asked first to remove the two definitely wrong options and then to select the correct option from the remaining two. This hypothesis was tested at the school level.

Hypothesis 3.5: There is no significant difference in the item's facility value, when the position of the most plausible distractor is altered, when students are asked to remove first the two definitely wrong options and then select the correct option from the remaining options. This hypothesis was aimed to test at the university level.

3.2.5.4 The Method of the Research:

Our hypothesis suggests that if the differences in the facility values of the MCQs, when the positions of the key answer and the most plausible distractor are rearranged is due to the 'optical' effect, then this difference will disappear or at least be reduced.

To examine if the differences in the facility values of the MCQs has an optical origin, the intention was to make two investigations; one at school level and one at university level.

i. The School Level:

a. Description of the Multiple Choice Test:

The test consists of ten chemistry multiple choice questions (Appendix 3). It was designed in two forms; A and B. Form (A) is a normal MCQs test in which pupil was asked to select the correct answer. In Form (**B**) the examinee needs to do two steps in order to answer the questions completely. Step one required the examinee to designate the two options which he thought were definitely wrong by peeling adhesive strips carrying the distractors from the test and throwing them away. The second step required the examinee to select the correct option from the remaining two. Our aim of Form (**B**) is to leave the key answer and the most plausible distractor in the test paper without the interference of the two 'wrong' distractors. Consequently, a pupil is able to compare between only two options to reach the correct or the best answer.

The options of the Form (**B**) test were written in 46 $\times$ 11.11mm labels. The reason behind this is to allow pupils to peel off the two definitely wrong options and discard them (Appendix 3).

Three versions of Form (**B**) were constructed. The only difference between the three versions of Form (**B**) was the position of the key answer which was altered from one version to another except question one, which was kept the same, to act as internal marker. The position of the key answer of Form (**A**) items and the items of the original version of Form (**B**) was the same in all test items.

b. The Marking Scheme:

To mark the test, one mark has been allocated for the correct answer of Form (**A**) because it consists of normal multiple choice questions (i.e. select the correct answer from the given options). However, a different marking scheme was used to score the questions of Form (**B**). The marking scheme was presented in Table 3.13 below.

Table 3.13: Number of marks allocated to each type of selection in Form (B) of the school chemistry multiple choice test

Type of Selection	Number of Marks
Selected the actually correct option as the correct one	3
Did not select actually correct option as the correct one or one of the definitely wrong options	2
Selected the actually correct option as one of the definitely wrong options	0

Let us take an example from chemistry test we applied in school to show how the above marking scheme was work.

- Which one of the following elements has an electron arrangement of 2, 8, 2?
    a. oxygen.

    b. calcium.

    c. magnesium.

    d. fluorine.

The correct answer is option **c.** We will take five pupils as examples with different types of selection (Table 3.14).

Table 3.14: Number of marks allocated to each pupil according to his selection in Form (B) of the school chemistry multiple choice test

	Type of Selection (Options)		
Pupil	**Correct Answer**	**Two Definitely Wrong Options**	**Marks**
1	c	a, b	3
2	d	a, b	2
3	a	c, d	0
4	c	b, d	3
5	b	a, d	2

c. The Sample:

The sample was one hundred and ninety (190) fourth year secondary pupils. They were selected from four high schools in the Central Region of Scotland. Fifty (50) pupils answered Form (A), whereas the number of pupils who answered the three versions of Form (B) were as follows:

- 48 pupils answered the original version.
- 45 pupils answered version one.
- 47 pupils answered version two.

ii. The University Level:

a. Description of the Multiple Choice Test:

At the university level, the plan was to apply one test to find out more evidence if the differences in the facility values of the MCQs is something optical. The test was based on the Evidence for Evolution course. The test was one, which was used to investigate the effect of changing the position of the most plausible distractor on the facility value of the multiple choice questions (Appendix 1.1). It consisted of thirty (30) multiple choice questions. The proposal was to design three versions of the test and the only difference between them was the position of the most plausible distractor relative to the key answer. In the original version the most plausible distractor was placed next to the key answer, in version one the most plausible distractor was placed one position away from the key answer and in version two the most plausible distractor was placed furthest away from the key answer.

The intention was to administer the test by the computer based assessment package: 'Tripartite Interactive Assessment Delivery System' (**TRIADS**). This package, like **'Question Mark'**, allows for randomization (For more details about Question Mark and **TRIADS** packages see Chapter Two, section 2.3.4).

The plan was to use the same system of answering as in the school test. It consisted of two steps. Step one required the examinee to designate the two options which he thought were definitely wrong by clicking on them (Figure 3.1). They automatically faded from the computer screen. The purpose of this was to bring up the key answer and the most plausible distractor as close as possible without the interference of the already eliminated distractors. Therefore, the student was able to compare between them more effectively to reach the correct or the best answer.

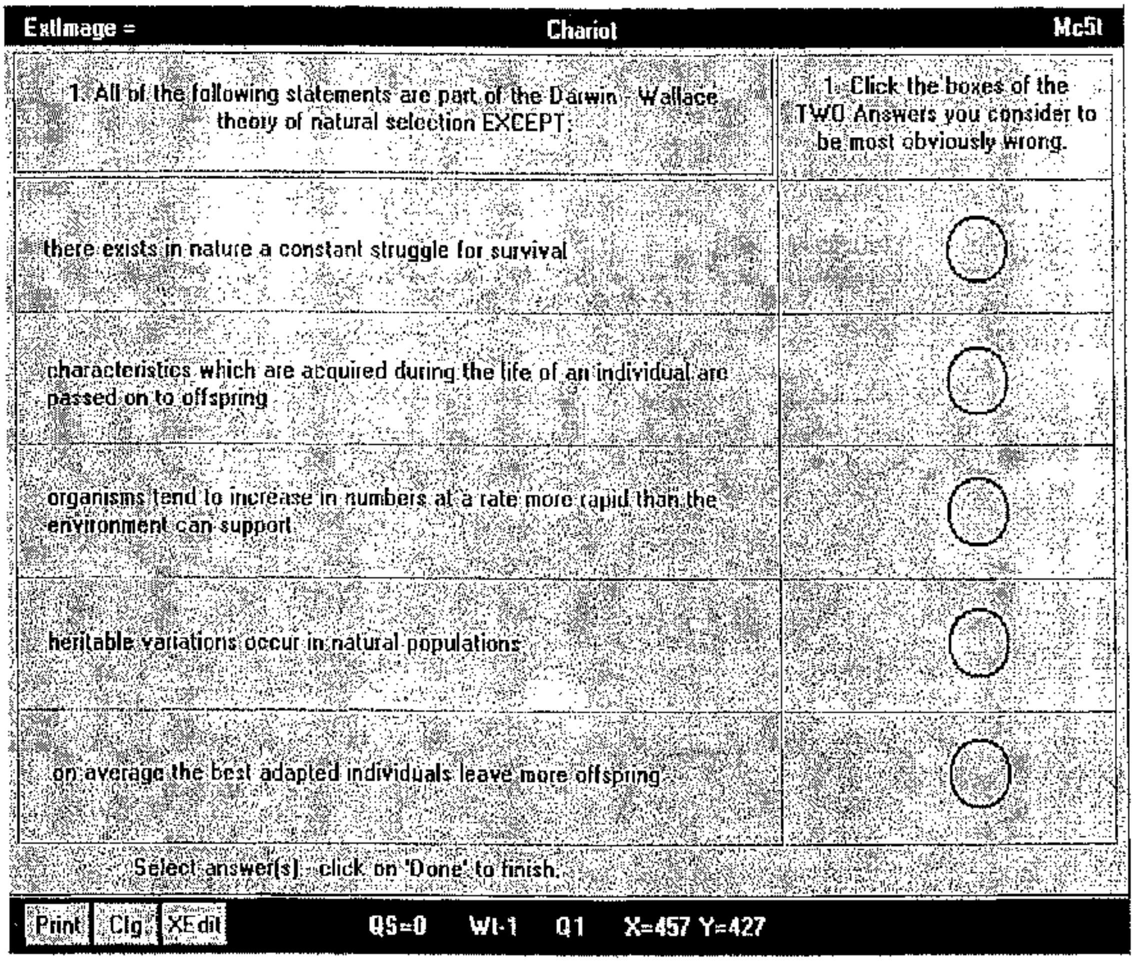


Figure 3.1: The view of the computer screen when students were asked to do the first step of answering the question

In the second step, the examinee was required to select the correct option from the remaining options (Figure 3.2).

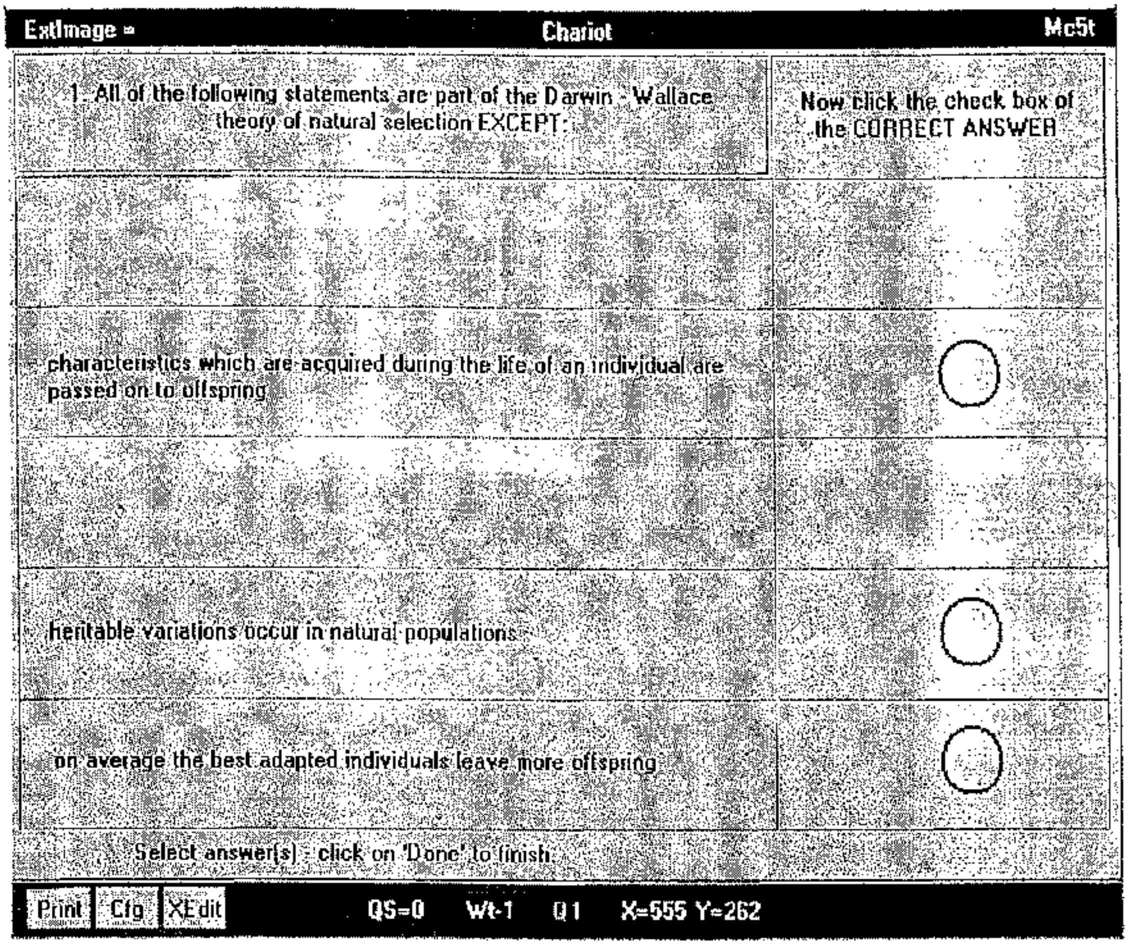


Figure 3.2: The view of the computer screen when students were asked to do the second step of answering the question

After the student did the two steps, the computer asked him if he was sure about the answer. If he said 'Yes', then a new question appeared. However, if he said 'No', there is an opportunity for the student to select another option before quitting (Figure 3.3).

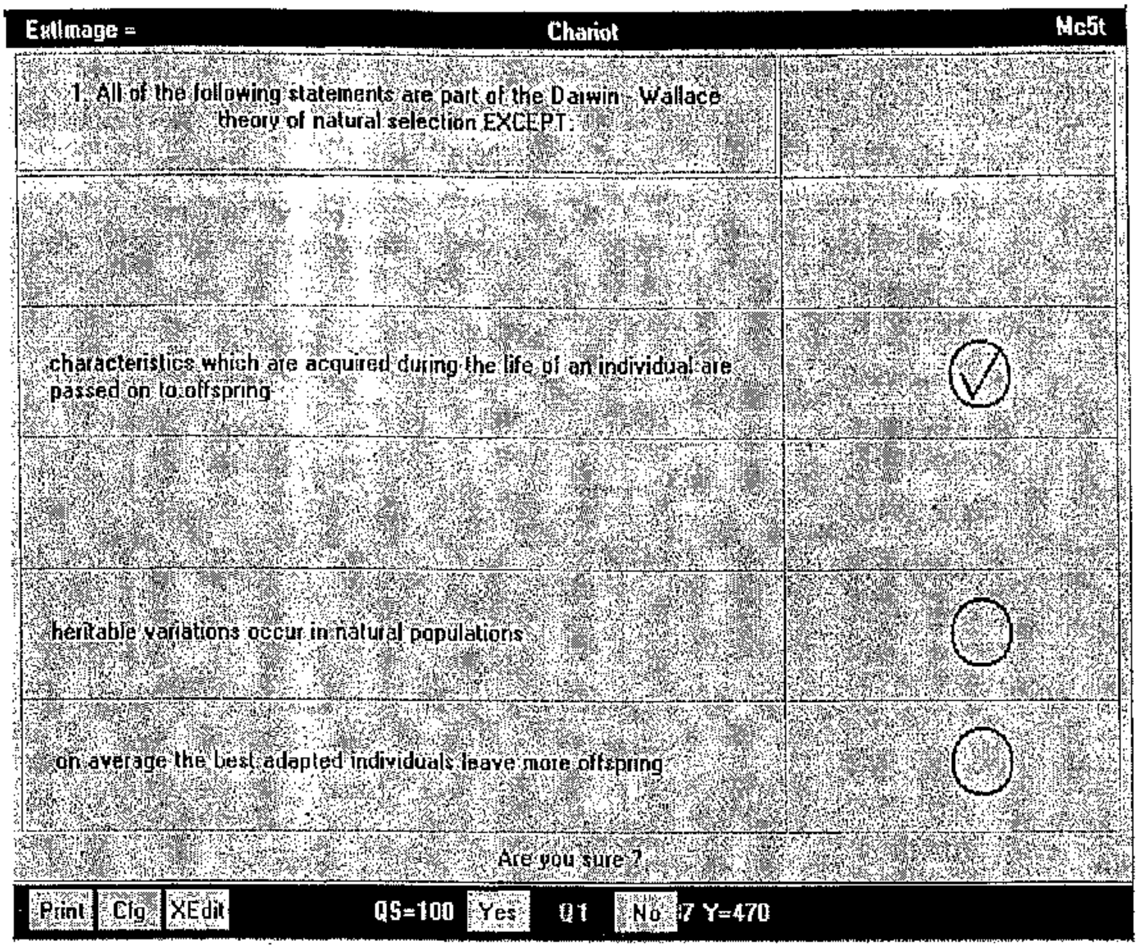


Figure 3.3: The view of the computer screen after the students selected the correct answer

Unfortunately, the computer server at the Institute of Biomedical and Life Science – University of Glasgow, the place where the test supposed to be applied, broke down. It happened at a crucial time for the researcher and students. It occurred two weeks before the students finished their study. Consequently, the researcher would have to wait for another year in order to apply the test again, which he cannot do due to the limit on his study period (i.e. scholarship).

Contact was made with the people who were in charge of the computer server to ask how long it would take to fix or replace the server. The reply was that it would take longer than was possible for our project. It is really sad to lose this part of the research. However, the hope is that a new researcher will come and continue this work using the **'TRIADS'** computer package.

3.2.6 The Results of Examining the "Optical" Explanation:

The factor, which we think may be partially the cause of the differences in the facility values of the MCQs when the positions of the key answer and the most plausible distractor have been rearranged, is something optical. Let us see if the results support our hypothesis.

3.2.6.1 The Mean Score of the Two Forms of the Test:

As it was mentioned previously that, in order to examine if the differences in the facility values of the MCQs when the position of the key answer has been rearranged is something optical, two forms of the multiple choice test were applied to S4 pupils in chemistry. These are Form (**A**) which asked pupils to answer the questions by drawing a circle around the correct answer and Form (**B**) which asked pupils first to designate the two options thought to be definitely wrong by removing them from the test paper and then to draw a circle around the option thought to be the correct one.

The mean score of the original version of Form (**B**) was used to compare with the mean score of Form (A) (Table 3.15). This is due to the fact that the two forms were identical in terms of the positions of the options.

Table 3.15: The mean score and the standard deviation of Form (A) and the original version of Form (B) of the school chemistry multiple choice test

	FORM A	**FORM B (Original Version)**
Number of Pupils	50	48
Mean Score	21.4	27.0
SD	6.05	4.31

Note:

The final score of each pupil of Form (A) was multiplied by 3 for comparison purpose.

As can be seen from Table 3.15, the mean score of those pupils who did the original version of Form (**B**) is higher than the mean score of those pupils who did the Form (**A**). The difference in the mean scores between the two groups was found to be statistically significant at 1% levels when the t - test was applied (Degree of freedom = 96, two tailed test). As a result of this, hypothesis **3.3** which states that *'there is no significant difference in the item's facility value when pupils/students were asked to answer MCQs by two different methods; the normal method (select the correct option) and by the* *method of removing the two definitely wrong options and then select the correct one from the remaining options'* can be rejected.

3.2.6.2 Pupils' Performances in Each Item of the Two Forms of the Test:

How did pupils perform in each item of the two forms of the test? To find an answer to this question, the facility value of each item of Form (**A**) and the original version of Form (**B**) was calculated and presented in the Table 3.16 below.

Table 3.16: The facility value of each item of Form (A) and the original version of Form (B) of the school chemistry multiple choice test

	Facility Value	
Item Number	**FORM A**	**FORM B (Original Version)**
1	0.94	0.94
2	0.68	0.92
3	0.90	0.90
4	0.64	0.90
5	0.76	0.94
6	0.76	0.94
7	0.80	0.96
8	0.38	0.75
9	0.48	0.81
10	0.78	0.92

From the table above, the facility value of item one which was kept without any change is the same in both tests. On the other hand, the facility value of all items, which were exposed to change, is higher in the original version of Form (**B**) than the facility value of the same items of Form (**A**) except item number three. In Item number three, the facility value is the same in Form (**A**) and the original version of Form (**B**). The difference in the facility value between the two forms was found, in some items, to be more than 20% (Items **2**, **4**, **8**, **9**). To find out if this difference is statically significant or not, Kellett's (1978) method was used (Appendix 1.3). The results of this calculation showed that the difference in the facility was statistically significant at the 1% level for items **2, 4, 5, 6, 7, 8, 9, 10**. To remind the reader, the only difference between these identical tests (Form (**A**) and the original version of Form (**B**)) was the method by which the pupils tackled them.

3.2.6.3 The Mean Score of the Three Versions of Form (B):

As we know from section 3.2.5.4 that three versions of Form (B) were designed. The only difference between them was the position of the key answer. Pupils' mean scores in the three versions were calculated, presented and compared in Table 3.17 below.

Table 3.17: The mean score and the standard deviation of the three versions of Form (B) of the school chemistry multiple choice test

	Original Version	Version one	Version Two
Mean Score	27.0	26.2	27.1
SD	4.31	3.97	3.24

It can be seen from Table 3.17 that there is almost no difference in the mean score between the original version and version two. However, there is a slight difference between these two versions and version one. To find out if this difference is statistically significant or not, *t* - test was applied. The results showed that the difference in the mean score between the three versions is not significant.

Now let us move to look at the facility value of each item of the three versions (Table 3.18).

Table 3.18: The facility value of each item of the three versions of Form (B) of the school chemistry multiple choice test

	Facility Value		
Item Number	**Original Version**	**Version One**	**Version Two**
1	0.94	0.95	0.95
2	0.92	0.93	0.96
3	0.90	0.93	0.85
4	0.90	0.87	0.89
5	0.94	0.86	0.94
6	0.94	0.87	0.98
7	0.96	0.97	0.91
8	0.75	0.61	0.77
9	0.81	0.90	0.82
10	0.92	0.83	0.91

As we can see from Table 3.18 that the difference in the facility value between most of the items of the three versions is less than 10% except item six (between version one and version two) and item eight (between version one and the original version and version two).

To find out if the difference in the facility value of items six (6) and eight (8), which have the difference from one version to another by 10%, is significant or not, the method designed by Kellett (1978) was used (Appendix 1.3). The results of this calculation are presented in Table 3.19.

Table 3.19: The level of significance of items six and eight of Form (B) of the school chemistry multiple choice test

	Items	
Level of Significance	**Original Version and Version One**	**Version One and Version Two**
5%	8	6, 8
1%	-	8

From the table above, the difference in the facility value of item six which occurs between version one and version two is statistically significant at 5% level. However, for item eight it was found that the difference in the facility value between the original version and version one is statistically significant at 5% level and between version one and version two is statistically significant at 5% and 1% levels. Consequently, the null hypothesis (hypothesis 3.4); that is '*there is no significant difference in the item's facility value, when the position of the key answer is altered, if pupils are asked first to remove the two definitely wrong options and then to select the correct option from the remaining two*' is accepted for all items except item six and eight.

However, we can look to the 'optical' factor above from the angle of Information Processing Model. Information Processing Model is an attempt to suggest mechanisms for learning arising from a number of psychological schools (Johnstone, 2000a). In this model, the perception (how we take a first view of something) is controlled by what we already know and believe (Johnstone, 2000b). The filtered material is admitted into the conscious part of our mind (Working Space) for further processing. The working space is limited and we can consciously handle only a limited amount of information in a given time. If we try to manipulate too much at once, learning become faulty or not take place at all, because we just overload and shut down. Then, if we decide to store any information, we look for clear attachments in our Long Term Memory on which to fix our new knowledge or experience.

Return back to the main point, the proximity effect upon the facility values of MCQs was almost eliminated when pupils were asked to tackle the questions in two steps:

- Eliminated the two options you think are definitely wrong.
- Now compare the remaining options and decide which is correct.

The elimination by physical means had the effect of removing the 'noise' of the question and simplifying the final choice. This had the effect of achieving stable facility values regardless of the initial relative position of the key answer and strong distractor. This strongly suggests that some 'optical' effect, or just a lowering of processing demand, leads to stable facility values.

Finally, what is to be said that, in the current research only a few items (10) were used to investigate the 'optical' factor which might be behind the differences in the facility values of MCQs. Therefore, it is recommended that this should be examined with larger number of items to support the results of the current research.

CHAPTER THREE: FIELD RESEARCH AND RESULTS (1) - PART (B)

CHAPTER THREE: FIELD RESEARCH AND RESULTS (1) - PART (B)

3.3 Fixed Response: Structural Communication Grids (SCGs)

3.3.1 Introduction:

The results of part (A) of the current chapter showed that a small amount of editing can bring a difference in the facility value of multiple choice questions by 10% or more. However, in Chapter Two (section 2.2.1.5) a fixed response method known as structural communication grids (SCGs) was discussed and the results of some previous studies on this method suggested that it is powerful for assessing students' performance. It seems to be that this method can overcome some problems that multiple choice questions have. The **question** which could be highlighted here is '*Does the facility value of the grid questions also suffer, like MCQs, from minor editing?*'. The following paragraphs will answer this question.

3.3.2 The School Level:

3.3.2.1 The Aims of the Research:

The aims of this test are to investigate whether:

- SCGs suffer changes in facility value from recasting the positions of the options in the grid.
- Pupils' ranking is changed when two different scoring methods are used to score SCGs.

3.3.2.2 The Research Hypotheses:

Two null hypotheses were tested at school level in this part of the research:

Hypothesis 3.6: There is no significant difference in the facility value of the question (sub – question) when only the positions of the options are altered within the grid.

Hypothesis 3.7: There is no correlation between pupils' rank order when two scoring methods are used to score the grid questions.

3.3.2.3 The Marking Scheme:

The researcher marked the test himself. For 1997/98 academic year, the researcher marked the test using the method suggested by Egan to score the SCGs (Full details of the method suggested by Egan can be found in Chapter Two, section 2.2.1.5).

In the second academic year (1998/99), the researcher used the Scottish Qualifications Authority scoring method and the method suggested by Egan to score the SCGs. The reason for this is to compare pupils' ranking in the two methods. To remind the reader, in the Scottish Qualifications Authority scoring method one mark has been allocated for each closed question and one mark has been allocated for each correct selection in an open question (For more details on the Scottish Qualifications Authority scoring method, see Chapter Two, section 2.2.1.5).

3.3.2.4 The Method of Research and Results:

i. Description of the Test:

The test consists of eight 3×2 **or** 2×3 grid questions. Most of the questions were divided into two or three sub - questions except question one. The test was applied to S4 Standard Grade pupils (Age 15-16) in chemistry. Two versions of the test have been designed (Appendix 4.1). There is no change in any word or in the chemistry data in the two versions except that the positions of the options were randomly altered within the grid.

Examples of the School SCGs Test:

Example One:

Version One:

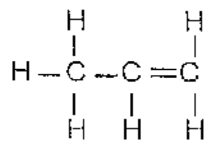
* Hydrocarbon compounds are obtained from crude oil.

Butene A	Propane B	Hexene C
Pentane D	Pentene E	Propene F

(a) Select the TWO hydrocarbons which are Alkanes?

A	B	C
D	E	F

(b) Select the hydrocarbons which has the following structural formula?



A	B	C
D	E	F

Version Two:

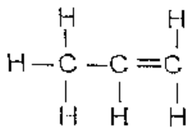
* Hydrocarbon compounds are obtained from crude oil.

Butene A	Propene B	Hexene C
Propane D	Pentene E	Pentane F

(a) Select the TWO hydrocarbons which are Alkanes?

A	B	C
D	E	F

(b) Select the hydrocarbons which has the following structural formula?



A	B	C
D	E	F

Example Two:

Version One:

* Chemical reactions make new substances.

Ice melting A	Neutralising acid **B**	Iron rusting **C**
Breaking glass **D**	Methane burning **E**	Polymerising ethene **F**

(a) Select the TWO changes which are not examples of chemical reactions?

A	B	C
D	E	F

(b) Select the chemical reaction(s) which would produce water as one of the products?

A	B	C
D	E	F

Version Two:

* Chemical reactions make new substances.

Ice melting A	Breaking glass **B**	Iron rusting **C**
Neutralising acid **D**	Methane burning E	Polymerising ethene **F**

(a) Select the TWO changes which are not examples of chemical reactions?

A	B	C
D	E	F

(b) Select the chemical reaction(s) which would produce water as one of the products?

A	B	C
D	E	F

ii. First Academic Year (1997/98):

a. The Sample:

In 1997/98 academic year, the test was applied in two secondary schools in Glasgow area. The number of pupils who sat the test was 60 pupils (Table 3.20).

Table 3.20: Number of pupils who answered the two versions of the SCGs from the two high schools (1997/98)

	Number of Pupils	
	Version One	**Version Two**
School One	15	16
School Two	14	15
Total	**29**	**31**

b. The Results:

The facility value of each sub - question of the two versions of the test is presented in Table 3.21.

Table 3.21: The facility value of each sub - question of the two versions of the Standard - Grade chemistry SCGs (1997/98)

	Facility Value	
Sub-Question	**Version One**	**Version Two**
1.a	0.86	0.83
2.a	0.97	1.00
2.b	0.46	0.73
3.a	0.97	1.00
3.b	0.97	0.66
4.a	0.79	0.81
4.b	0.39	0.39
5.a	0.66	0.66
5.b	0.75	0.85
5.c	0.74	0.81
6.a	0.88	0.89
6.b	0.46	0.57
7.a	0.63	0.76
7.b	0.40	0.42
8.a	0.38	0.09
8.b	0.38	0.36

From the table, it is clearly shown that changing the positions of the options within the grid can bring a difference in the facility value of the grid questions by 10% and more. The following are examples of the sub - questions which have a difference in the facility value by 10% or more.

The facility value of the version one of sub - question **2.b** is 0.46. However, when the positions of the options were altered in version two the facility value increased to 0.73.

Sub - question **3.b** is another example. In this sub - question, the facility value decreased from 0.97 in the version one to 0.66 in version two.

Sub - question **5.b** too suffered from changing the positions of the options within the grid. The facility value of version one of this sub – question is 0.75. However, when the positions of the options were altered in version two, the facility value increased to 0.85.

The fourth example is sub – question **6.b**. In this sub – question, the facility value of version one is 0.46. When the positions of the options were altered in version two the facility value increased to 0.57.

The fifth example is sub – question **7.a**. The facility value of version one is 0.63, whereas the facility value of version two is 0.76.

The final example of those sub – questions, which have the difference in their facility value by 10% or more, is sub - question **8.a**. This sub - question has the biggest difference in facility value amongst other sub – questions of the test. The difference in the facility value between the version one and version two is 29%

Nevertheless, are these differences statistically significant or not? To find out which of these differences in the facility value are significant or not, Kellett's (1978) method was applied. Table 3.22 shows the level of significance of some sub - questions.

Table 3.22: The level of significant of some sub- questions of Standard - Grade chemistry SCGs (1997/98)

Level of Significance	Sub - Questions
5%	2.b, 3.b, 8a
1%	2.b, 3.b, 8a

From the table, three out of seven of those sub – questions which have a difference in facility value between the version one and version two are statistically significant at 5% As a result of this, hypothesis **3.6** can be rejected for sub – questions **2.b**, **3.b**, and **8.a**.

iii. Second Academic Year (1998/99):

a. Introduction:

In the second academic year (1998/99), the same test was applied again to a larger sample to confirm the findings of the previous academic year (1997/98).

b. The Sample:

The number of pupils who did the test this time was one hundred and seventy three (173) pupils from four high schools in the Central Region of Scotland (Table 3.23). The test was administered to pupils by their class teachers under the normal exam conditions.

Table 3.23: Number of pupils who answered the two versions of the SCGs test from the four high schools (1998/99)

	Number of Pupils	
School	**Version One**	**Version Two**
School One	25	19
School Two	16	9
School Three	23	30
School Four	25	26
Total	**89**	**84**

c. The Results:

The results of the first academic year (1997/98) showed that changing the positions of the options within the grid bring, to some extent, a significant effect on the facility value of some sub – questions. We would like now to know if the results of the second academic year (1998/99) match the results of the first academic year.

The facility value of each sub – question was calculated and tabulated in Table 3.24 below.

Table 3.24: The facility value of each sub - question of the two versions of the Standard - Grade chemistry SCGs (1998/1999)

	Facility Value	
Sub-Question	**Version One**	**Version Two**
1.a	0.66	0.65
2.a	0.89	0.95
2.b	0.29	0.37
3.a	0.98	0.95
3.b	0.76	0.74
4.a	0.71	0.68
4.b	0.29	0.23
5.a	0.70	0.80
5.b	0.87	0.68
5.c	0.76	0.62
6.a	0.91	0.88
6.b	0.35	0.33
7.a	0.46	0.43
7.b	0.43	0.44
8.a	0.12	0.03
8.b	0.45	0.24

The same results we came up with in 1997/98 academic year were also found in the second academic year (1998/99) in terms of the effect of changing the positions of the options within the grid in the facility value of the question. However, the number of items which have the difference in their facility value from one version to another by 10% or more was smaller compared with 1997/98 academic year. In 1998/99 academic year, the number of sub – questions which have the difference in the facility value by 10% or more is only four.

The facility value of sub - question **1.a**, is almost the same in the two versions of the test. This sub – question as was mentioned earlier was not exposed to any change to act as an internal marker to check the groups' sampling. Therefore, if there is a difference in the facility value of some sub – questions, which were subjected to changes, then we could assume that it is due to these changes.

The first example is sub - question **5.a**. The difference in the facility value between version one and version two is 10% (F.V of version one = 0.70 and of version two = 0.80).

The second example is sub - question **5.b**. The difference in the facility value between the two versions is by 19% (F.V of version one = 0.87 and of version two = 0.68).

The third example is sub - question **5.c**. In this sub - question, the difference in the facility value between the two versions is by 10% or more. The facility value of version one is 0.76. However, when the positions of the options were altered within the grid in version two, the facility value decreased to 0.62.

The final example is the last sub - question of the test (**8.b**). In this sub – question, the difference in the facility value between the original version and version one is 21% (F.V of the original version = 0.45 and for version one = 0.24).

Now we would like to find out which of the above differences in the facility value, which was found for some sub – questions of Table 3.24, are statistically significant (Table 3.25). Kellett's (1978) method was used to calculate the level of significance of these differences (Appendix 1.3). Table 3.25 shows the sub - questions which have statistically significant difference in their facility value from one version to another.

Table 3.25: The level of significance of some sub- questions of Standard - Grade chemistry SCGs (1998/1999)

Level of Significance	Sub - Questions
5%	5.a, 5.b, 5.c, 8.b
1%	5.b, 5.c, 8.b

From the table, hypothesis **3.6**; that is '*there is no significant difference in the facility value of the question (sub – question) when only the positions of the options are altered within the grid*' can be rejected for sub - questions **5.a, 5.b, 5.c** and **8.b.**

What about if we look to each grid as whole? In other words, will there be any difference in the facility value of the grid if it is treated as one question? Table 3.26 shows the facility value of each question of the two versions of the Standard – Grade chemistry grid questions.

Table 3.26: The facility value of each question of the two versions of the Standard - Grade chemistry SCGs (1998/1999)

	Facility Value	
Question	**Version One**	**Version Two**
1	0.66	0.65
2	0.59	0.66
3	0.87	0.85
4	0.50	0.46
5	0.78	0.70
6	0.63	0.61
7	0.45	0.44
8	0.29	0.14

It is noted from the table above that the difference in the facility value of each question between the two versions is less than 10% except for question eight. That is, if we are taking each question as a whole the difference in the facility value of some sub – questions which was found above has disappeared (Table 3.24). The difference in the facility value of question eight was found statistically significant at 1% level when Kellett's (1978) method was applied. Therefore, hypothesis **3.6** can be rejected for question number eight (**8**).

Moving to another point of the school grid questions. We would like to find out the overall performance of the pupils in each version of the test. This can easily be done by calculating the average of the facility values of each version of the test (Table 3.27).

Table 3.27: The average of the facility values of the two versions of the Standard - Grade chemistry grid questions (1998/99)

	Test Version	
	Version One	**Version Two**
Average of the Facility Values	0.61	0.56

From the table, we can see that the difference in the average of the facility values between the two versions is by 5%. However, this difference was found statistically not significant when the t – test was applied.

Now let us assume that we would like to design a test consisting of sixteen (16) sub – questions. These sub – questions are randomly selected from the two versions of the test to be used in three occasions. Each occasion consists of a different sixteen sub – questions For example, on occasion one we selected sub – questions **1.a**, **2.a**, **2.b**, **4.a**, **4.b**, **5.a**, **5.b**, **5.c**, **8.a** and **8.b** from version one, sub – questions **3.a**, **3.b**, **6.a**, **6.b**, **7.a** and **7.b** from version two (Table 3.28).

The aim of this is to examine pupils' performances in the test as a whole when the sub – questions are selected from the two versions.

Table 3.28: The sub – questions of the three occasions of the school SCGs which are selected to check pupils' performances in the test as whole

	Occasion One		Occasion Two		Occasion Three	
Sub Question	**Version One**	**Version Two**	**Version One**	**Version Two**	**Version One**	**Version Two**
1.a	✓			✓	✓	
2.a	✓			✓		✓
2.b	✓			✓		✓
3.a		✓	✓		✓	
3.b		✓	✓		✓	
4.a	✓			✓		✓
4.b	✓			✓		✓
5.a	✓			✓	✓	
5.b	✓			✓	✓	
5.c	✓			✓	✓	
6.a		✓	✓			✓
6.b		✓	✓			✓
7.a		✓	✓		✓	
7.b		✓	✓		✓	
8.a	✓			✓		✓
8.b	✓			✓		✓

Note:-

✓ = *sub – questions are selected to include in each occasion.*

To find out the average of the facility values of the test on each occasion, the facility value of all the sub - questions is added together and then divided by 16 (the total sub – questions in the test).

The results of this calculation showed that the average of the facility values of the test as a whole is not too much different on the three occasions. It has been found that the average of the facility values of the test as a whole for occasion one is 0.59, whereas for occasions two and three the values are 0.57 and 0.58 respectively. The difference in the average of the facility values between the three occasions is about 5%, which is acceptable.

iv. Two Scoring Methods to Score SCGs and Testing the Hypothesis:

One aim of applying SCGs at school level is to compare pupils' ranking in two methods of scoring; (a) Scottish Qualifications Authority and (b) the method suggested by Egan.

Egan's method of scoring allocates scores to pupils according the following formula:

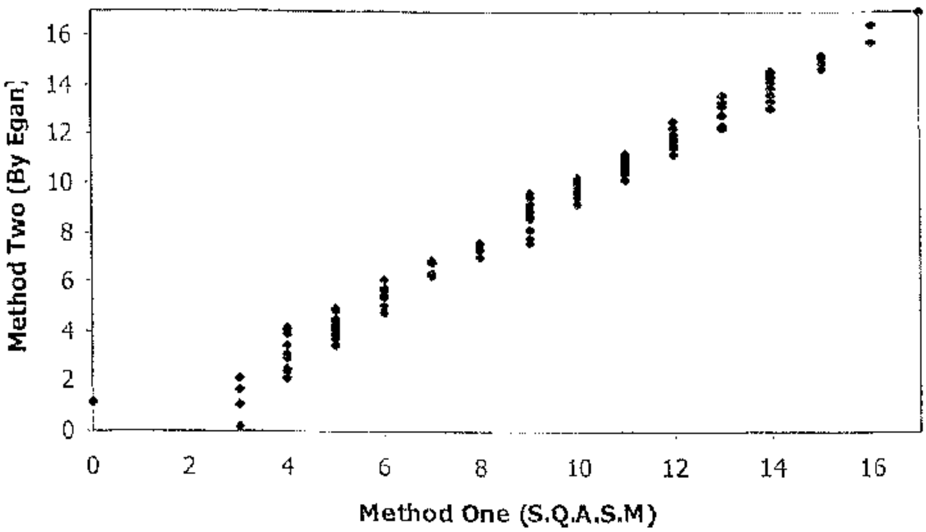
Coefficient of Confusion (Scores) =

$$\frac{\text{No. of relevant } (\checkmark) \text{ responses chosen}}{\text{No. of relevant } (\checkmark) \text{ responses available}} - \frac{\text{No. of irrelevant } (\times) \text{ responses chosen}}{\text{No. of irrelevant } (\times) \text{ responses available}}$$

The resultant values of the "coefficient of confusion" range from +1, through 0 to -1, which then can be added or multiplied by any number.

Before looking at the rank order correlation between the two methods of scoring, it is worthwhile to find out the Pearson Product - Moment Correlation Coefficient (r) between the two methods. Pupils' scores in both methods were plotted against each other (Figure 3.4).

Figure 3.4: A scatter plot between pupils' scores in Scottish Qualifications Authority Scoring Method and their scores in the method suggested by Egan



Note:-

Method One: *Scottish Qualifications Authority.*

Method Two: *Method suggested by Egan.*

The value of Pearson Product - Moment Correlation Coefficient (r) showed that there is a highly positive correlation ($r = 0.993$, Degrees of freedom 171/two – tailed) between the Scottish Qualifications Authority scoring method and the method suggested by Egan.

Looking at the above figure, we can see that pupils who got a high score in the Scottish Qualifications Authority scoring method, in general, also got a high score in the method suggested by Egan. However, there are some exceptions. For example, one pupil got zero when the Scottish Qualifications Authority method of scoring was used. On the other hand, when the method suggested by Egan was used the same pupil got 1.12. Another pupil got 3 marks when the Scottish Qualifications Authority scoring method was used but he got 0.21 when the method suggested by Egan was used.

This is due to the fact that in the latter method (the method suggested by Egan) there is a possibility that pupil can get credit for partial knowledge and penalization of the wrong selection(s) but this is not the case in the former method (Scottish Qualifications

Authority).

Now we would like to come back to hypothesis 3.7, which states that '*there is no correlation between pupils' rank order when two scoring methods are used to score the grid questions*'. To test this hypothesis, Spearman Rank Order Correlation Coefficient (*rho*) was applied. The result of the correlation carried out showed that there is a high positive correlation between pupils' ranking in the two methods (*rho* = 0.995, N = 173) Therefore, the hypothesis 3.7 can be rejected at 0.001 % level of significance.

However, looking at the ranking of pupils in the two methods (Appendix 4.2), we can see that the Scottish Qualifications Authority scoring method has more ties (i.e. many pupils in the same rank) than the method suggested by Egan. Moreover, the method which is suggested by Egan, discriminates better between pupils than the Scottish Qualifications Authority scoring method.

v. Summary of the Results of Applying SCGs at the School Level:

We can summarise the above results as follows:

- The facility value of the SCGs also suffers to some extent, like MCQs, from minor editing in the question. The editing which happened to SCGs, was changing the positions of the options within the grid.
- There is no difference in pupils' rank order between the Scottish Qualifications Authority scoring method and the method suggested by Egan.
- The difference in the average of the facility values of the two test versions was by 5%.
- The difference in the average of the facility values of the test as whole in three occasions (each occasion consists of 16 sub – questions from the two versions) was less than 5%

3.3.3 The University Level:

3.3.3.1 The Aim of the Research:

The aim of applying SCGs at the university level is to get further evidence on the effect of changing the positions of the options within the grid in the facility value of the question.

3.3.3.2 The Research Hypothesis:

One hypothesis was tested in this part.

Hypothesis 3.8: There is no significant difference in the facility value of question (sub – question) when the positions of options are altered within the grid at the university level.

3.3.3.3 The Marking Scheme:

The researcher used the method suggested by Egan (**Coefficient of Confusion**) to mark the selection part of the SCGs and used Mackenzie's third method (**Before and Next to**) to score the ordering of students' selections.

3.3.3.4 The Method of Research:

i. Description of the Test:

The test consists of six **8 × 2 or 2 × 8** grid questions. Most of the questions were divided into two or more sub - questions. Two sub - questions (**2.c** and **2.d**) asked students to put their selections in a logical order. The test was designed for first year university students in chemistry at the University of Glasgow. The researcher designed the test based on the objectives set up by the Chemistry Department of the University. The test was reviewed by the specialists in chemistry from the Centre for Science Education - University of Glasgow and by the course lecturer from the Chemistry Department to check out the chemistry data, the layout and to get further suggestions or comments.

After all the revision process took place, two versions of the test were designed (Appendix 4.3). There was no change in any word or in the chemistry data in the two versions except that the positions of the options were altered randomly within the grid.

It is important to notice that question one (consists of two sub – questions) was kept without changes in the two versions to use as internal markers to check if student samples were comparable

Examples of the University SCGs Test:

Example One:

Version One:

${}^{25}_{54}Mn$ **1**	${}^{17}_{35}Cl$ **2**	${}^{3}_{6}Li$ **3**	${}^{35}_{79}Br$ **4**
${}^{8}_{15}O$ **5**	${}^{5}_{10}B$ **6**	${}^{14}_{28}Si$ **7**	${}^{9}_{18}F$ **8**

a. Which box(es) contain(s) metals?

_____ _____ _____ _____ _____ _____ _____ _____

b. Which box(es) contain(s) metalloids?

_____ _____ _____ _____ _____ _____ _____ _____

c.i Select the box(es) contain(s) halogens ?

_____ _____ _____ _____ _____ _____ _____ _____

ii. Order your selections in **c.i** above according to their **increase** in ionization energy?

_____ _____ _____ _____ _____ _____ _____ _____

d.i Select the box(es) contain(s) elements in the second row of the periodic table?

_____ _____ _____ _____ _____ _____ _____ _____

ii. Order your selections in **d.i** above according to their **decrease** in electronegativivty?

_____ _____ _____ _____ _____ _____ _____ _____

Version Two:

$^{25}_{54}Mn$ 1	$^{17}_{35}Cl$ 2	$^{5}_{10}B$ 3	$^{35}_{79}Br$ 4
$^{3}_{5}Li$ 5	$^{8}_{15}O$ 6	$^{9}_{18}F$ 7	$^{14}_{28}Si$ 8

a. Which box(es) contain(s) metals?

____

b. Which box(es) contain(s) metalloids?

____

c.i Select the box(es) contain(s) halogens ?

____

ii. Order your selections in **c.i** above according to their **increase** in ionization energy?

____

d.i Select the box(es) contain(s) elements in the second row of the periodic table?

____

ii. Order your selections in **d.i** above according to their **decrease** in electronegativivty?

____

Example Two:

Version One:

$Ca^{2+} \rightarrow Ca$ 1	$Cr_2O_7^{2-}$ 2	$HClO_4$ 3	Cl_2 4
H_2O_2 5	ClO_3^- 6	CrO_4^{2-} 7	$Sn^{2+} \rightarrow Sn^{4+}$ 8

a. Which box(es) contain(s) chlorine (**Cl**) in the oxidation state of +5?

____ ____ ____ ____ ____ ____ ____ ____

b. Which box(es) contain(s) the chemical equation which shows oxidation?

____ ____ ____ ____ ____ ____ ____ ____

c. Which box(es) contain(s) chromium (**Cr**) in the oxidation state of +6?

____ ____ ____ ____ ____ ____ ____ ____

d. Which box(es) contain(s) oxygen (**O**) in the oxidation state of -2?

____ ____ ____ ____ ____ ____ ____ ____

Version Two:

$Ca^{2+} \rightarrow Ca$ 1	$Cr_2O_7^{2-}$ 2	$HClO_4$ 3	Cl_2 4
$Sn^{2+} \rightarrow Sn^{4+}$ 5	H_2O_2 6	CrO_4^{2-} 7	ClO_3^- 8

a. Which box(es) contain(s) chlorine (**Cl**) in the oxidation state of +5?

____ ____ ____ ____ ____ ____ ____ ____

b. Which box(es) contain(s) the chemical equation which shows oxidation?

____ ____ ____ ____ ____ ____ ____ ____

c. Which box(es) contain(s) chromium (**Cr**) in the oxidation state of +6?

____ ____ ____ ____ ____ ____ ____ ____

d. Which box(es) contain(s) oxygen (**O**) in the oxidation state of -2?

____ ____ ____ ____ ____ ____ ____ ____

ii. The Sample:

Ninety six (96) first year university students completed the test. Forty seven (47) students answered version one and forty nine (49) answered version two.

3.3.3.5 The Results of Applying SCGs at the University Level:

Just to remind the reader before presenting the results of the grid questions at the university level, Egan's method of scoring was used to score the selection part, whereas Mackenzie's third method of scoring (**Before and Next to**) was used to score the logical order part.

The facility value of each sub – question of the two versions was calculated and tabulated in Table 3.29 below.

Table 3.29: The facility value of each sub - question of the two versions of university chemistry one SCGs

	Facility Value	
Sub – Question	**Version One**	**Version Two**
1.a	0.64	0.65
1.b	0.41	0.43
2.a	0.95	0.93
2.b	0.67	0.68
2.c.i	0.97	0.94
2.c.ii	0.64	0.63
2.d.i	0.73	0.81
2.d.ii	0.69	0.65
3.a	0.61	0.59
3.b	0.88	0.93
3.c	0.76	0.70
3.d	0.80	0.85
3.e	0.87	0.80
4.a	0.90	0.96
4.b	0.92	0.90
4.c	0.85	0.82
4.d	0.60	0.68
5.a	0.58	0.63
5.b	0.34	0.58
5.c	0.65	0.69
5.d	0.61	0.72
5.e	0.33	0.41
6.a	0.48	0.51
6.b	0.35	0.54
6.c	0.25	0.24
6.d	0.70	0.66

A examination of Table 3.29 leads us to conclude that the facility value of the grid questions does not suffer greatly from changing the positions of the options within the grid at the university level. It was found that three sub – questions only have a difference in their facility value from one version to another by 10% or more.

The difference in the facility value of the two sub – questions of **question one**, which were used as internal markers, was less than 10 % between the two versions. This indicates that the samples were similar. If there is a difference in the facility value of those questions, which were exposed to any changes, the differences are unlikely to be by sampling.

Sub - question **5.b** has a big difference in the facility value (24%) between version one and version two (F.V of version one = 0.34 and of version two = 0.58).

The second example is sub – question **5.d.** This sub - question has a difference in the facility value between version one and version two of 11% (F.V of version one = 0.61, and of version two = 0.72).

The last example which is worth looking at is sub – question **6.b**. The difference in the facility value between version one and version two was found to be 19%. The facility value of version one is 0.35. However, when the positions of the options were altered the facility value increased highly and became 0.54.

After we look at the facility value of selection type of questions. We would like to see the facility value of the logical order type of questions (i.e. the two sub – questions of question two (**2.c.ii** and **2.d.ii**) which asked students to put their selections in a logical order). The difference in the facility value of these questions was found less than 10% between the two versions.

It is now time to find out if the difference in the facility value, which was found for sub – questions **5.b, 5.d** and **6.b**, is statistically significant or not. Kellett's (1978) method was applied to calculate the level of significance of these sub – questions (Table 3.30).

Table 3.30: The level of significance of some sub - questions of university chemistry one SCGs

Level of Significance	Sub - Questions
5%	5.b, 5.d, 6.b
1%	5.b, 6.b

From the table, hypothesis **3.8** (i.e. *there is no significant difference in the facility value of question (sub - question) when the positions of options are altered within the grid at the university level*) can be rejected for sub – questions **5.b 5.d** and **6.b**.

Again as we did for the grid questions at the school level, we would like to see if there will be any difference in the facility value of the whole grid (i.e. treat the all sub – questions of one grid as one question). The following table (Table 3.31) shows the facility value of each question of the two versions.

Table 3.31: The facility value of each question of the two versions of university chemistry one SCGs

	Facility Value	
Question	**Version One**	**Version Two**
1	0.53	0.54
2	0.78	0.77
3	0.78	0.77
4	0.82	0.84
5	0.50	0.61
6	0.45	0.49

It can be seen from table above that one question (**question five**) has a difference in the facility value between versions one and two of 10% or more. The rest of the questions have a difference in the facility value of less than 10%. The difference in the facility value of question five was found statically significant at 5% level. Consequently, hypothesis **3.8** can be rejected for question number five.

We have seen above that there is a difference in the facility value of three sub – questions when the positions of the options within the grid were changed. What about the overall performance of the students in each version of the test? To answer this question, the average of the facility values of each version was computed and compared in Table 3.32.

Table 3.32: The average of the facility values of the two versions of university chemistry one grid questions

	Test Version	
	Version One	Version two
Average of the Facility Values	0.66	0.69

We can see from the table above is that there is a slight difference between the average of the facility values of the two versions of the test. This difference was found statistically not significant when the t – test was used.

We now investigate the students' performance in the test as a whole if the test consists of a mixture of sub – questions from the two versions. We assume that the sub – questions of the two versions were scrambled. Then, twenty six sub – questions were randomly chosen for each occasion from the two versions (Table 3.33). The average of the facility values of test as whole in each occasion was calculated by added together the facility value of each sub - question and then divided by 26 (the total sub – questions in the test).

The purpose of this investigation is find out the performance of students in the test as a whole when the sub - question were randomly selected from the two versions.

Table 3.33: The sub – questions of the three occasions of university chemistry one SCGs which are selected to check students' performance in the test as whole

	Occasion One		Occasion Two		Occasion Three	
Sub Question	Version One	Version Two	Version One	Version Two	Version One	Version Two
1.a	✓			✓	✓	
1.b	✓			✓	✓	
2.a	✓			✓		✓
2.b	✓			✓		✓
2.c.i	✓			✓		✓
2.c.ii	✓			✓		✓
2.d.i	✓			✓		✓
2.d.ii	✓			✓		✓
3.a		✓	✓		✓	
3.b		✓	✓		✓	
3.c		✓	✓		✓	
3.d		✓	✓		✓	
3.e		✓	✓		✓	
4.a	✓			✓		✓
4.b	✓			✓		✓
4.c	✓			✓		✓
4.d	✓			✓		✓
5.a		✓	✓		✓	
5.b		✓	✓		✓	
5.c		✓	✓		✓	
5.d		✓	✓		✓	
5.e		✓	✓		✓	
6.a		✓	✓			✓
6.b		✓	✓			✓
6.c		✓	✓			✓
6.d		✓	✓			✓

Note:-

✓ *= sub – questions are selected to include in each occasion.*

The results of the above investigation reported that the difference in the average of the facility values between the three occasions is less than 5%. The average of the facility values of the first occasion (occasion one) is 0.69, whereas the average of the facility values of the second occasion (occasion two) is 0.66. Finally, the average of the facility values of the third occasion (occasion three) is 0.66.

3.3.3.6 Summary of Results of Applying SCGs at the University Level:

We can summarise the above results as follows:

- The facility value of the grid questions also suffers, to some extent, from changing the positions of the options within the grid at the university level.
- The difference in the average of the facility values of the two test versions was less than 5%.
- The difference in the average of the facility values of the test as whole in three occasions (each occasion consists of 26 sub – questions from the two versions) was less than 5%.

3.3.4 Checking the Strategies Pupils Opt to Answer the SCGs:

3.3.4.1 Introduction:

Following the application of the grid questions in schools, an attempt has been made to investigate the strategies pupils use to answer this type of question.

3.3.4.2 The Method of Research:

A recording process took place in three out of four schools. In these schools, the researcher got permission from the pupils and their principal and class teachers before recording to make sure that pupils felt confident while they answered the questions. The researcher took notes for every pupil to support the recording. The research was conducted in two steps:

Step One: Asking pupils to read the question and answer it aloud.

Step Two: The researcher asked the pupil to give reason(s) for his/her selection(s).

3.3.4.3 The Sample:

Seventeen pupils (17) from the four high schools, where the SCGs test were firstly applied, were asked to answer the SCGs aloud (Table 3.34). They were selected from the top and bottom groups of the classes. Eight pupils were from the bottom group, whereas nine pupils from the top group. Two criteria have been used to select these pupils; their scores in the SCGs test and their class teacher's estimate.

Table 3.34: Number of pupils who participated in answering SCGs aloud from the four high schools

	Number of Pupils	
School	Bottom Group	Top Group
School One	1	3
School Two	2	2
School Three	3	3
School Four	2	1
Total	8	9

3.3.4.4 The Results of Checking the Strategies Pupils Opt to Answer SCGs:

The following points summarise the main findings of how pupils answer the exam when it is delivered in a grid form.

❖ Most of the good pupils (top group) preferred the grid to be presented in **3** × **2** (the shape below).

A	B	C
D	E	F

When the researcher asked these pupils why, they stated four reasons.

- Easy to compare between three options at the same time.
- Pupils not need to go up and down to search for the correct answer(s) as when the grid is presented in **2** × **3**.
- The above shape appears to pupils that there are few options in the grid.
- The shape of the grid where pupils put their answers (shape below) is similar to the grid where the options were written. Consequently, this will help pupils to reduce the confusions and the mistakes that they may make while they are entering their answer(s).

A	B	C
D	E	F

On the other hand, most weak pupils (bottom group) preferred the grid to be presented in 2 ×3 (the shape below):

A	D
B	E
C	F

The reasons for their preferences as follows:

- Easy to compare between two options at the same time.
- Easy to see because it looks larger in size.

❖ As with multiple choice questions, the guessing process is also likely to happen in grid questions. To remind the reader, for four options MCQs a pupil tends to guess between the remaining two options after he has eliminated the two definitely wrong options.

To understand how this happens in the grid questions with six boxes, we will take question five of the school level test as an example (Appendix 4.1). To answer sub – question **5.c**, the pupil first eliminates the two options which he already chose as the correct answer(s) for the previous two sub – questions; **5.a** (Mercury and Sodium) and **5.b** (Helium). Then, he makes another elimination, randomly, for one of the remaining options to end with two options, as is required in the sub - question. In the grid questions, some pupils might think that when one or more options had been selected as the correct answer(s) in the previous sub – questions they cannot be used to answer the new sub – question based on the same grid.

In addition, it was found that some pupils guess without any hint from the question. Because they do not know the correct answer and there is a 17% chance for each option in a grid with six boxes; pupils make a random guess. The scoring system, which is used currently in the schools (Scottish Qualifications Authority), does not penalize guessing. Therefore, it is recommended that another scoring system such as the one suggested by Egan should be used to minimize pupils' guessing.

❖ In general, pupils (especially the weak pupils) preferred closed type questions rather than open type questions. Closed questions are those questions in which the assessor indicates to the pupils in the stem of the question how many selection(s) are required to answer the question. The second type of question (i.e. open questions) are called open because the assessor does not indicate in the stem how many correct selection(s) that pupil should make.

The reasons why most pupils preferred closed questions are as follows:

- Closed questions do not require a lot of thinking like open questions.
- In the closed questions, the assessor indicates to the pupils in the stem how many selection(s) are required. As a result of this, pupils can narrow and eliminate some options and may reach the correct answers by guessing.
- There is less possibility for losing marks in closed questions compared to open questions. In open questions, the pupil does not know how many selection(s) are required to answer the question, and hence the pupil has more chance of marking mistakes.

  However, a small number of pupils (good pupils) preferred open questions because they see this type of questions as challenging.

❖ Finally, all pupils preferred the grid type questions to any other types of questions for the following reasons:

  - No need to write in the grid questions as opposed to open response (extended answer) questions. Pupils only need to select from the given options.
  - There is a possibility for guessing in closed questions, when pupils do not know the correct answer(s). They may get some marks by blind guessing in the closed questions.

3.3.5 Computer Program for SCGs:

3.3.5.1 Introduction:

A decision was made to design a computer program to provide shells in a grid form (CD attached to Thesis at the end). It was mainly designed:

- As a tool for pupils' self – assessment.
- To speed up the marking process of SCGs.

3.3.5.2 Description of the Program:

The program was designed to work in any personal computer (PC) which has Windows 95 or 97. The program was designed by Allan Lucas who was an undergraduate student in the Department of Computing Science – University of Glasgow. Here, the researcher would like to thank Allan Lucas and the Department of Computing Science for their help and assistance in designing the program.

A computer manual was written showing how the program can be installed and used (Appendix 5.1). The manual and the CD were given to somebody who is not expert in computing to install it in his computer. The aim of this is to make sure that everything is clear and understandable and could be done by an inexpert person.

The program was designed to score student's selection(s) according to the method suggested by Egan (Coefficient of Confusion) which can be calculated by the following formula:

Coefficent of Confusion (Scores) =

$$\frac{\text{No. of relevant (✓) responses chosen}}{\text{No. of relevant (✓) responses available}} - \frac{\text{No. of irrelevant (✗) responses chosen}}{\text{No. of irrelevant (✗) responses available}}$$

For scoring the logical order of student's selections, Mackenize's third method of scoring (**Before and Next to**) was used (for more information about the two methods, see Chapter Two, section 2.2.1.5).

The program consisted of three main components; Create Exam, Sit Exam and Edit Exam.

i. Create Exam Component:

In the Create Exam component (Figure 3.5, page 127), the examiner constructs the exam using any number of grid boxes up to 12. The exam's name and the file name (where the results will be stored) should be specified in this component. Three icons are found at the bottom of the component. These are **Create Intro** (to write an introductory paragraphs for the test as a whole or for each individual question), **Save Exam** (to save the exam in the hardware) and **Next Button** (for writing another question).

As was pointed in Chapter Two (section 2.2.1.5) SCGs can go beyond selection to ask students to put their selections into a logical order. This program allows the assessor to set up these questions and score them easily. In order to do so, the assessor should determine in the Create Exam component, while he is constructing the exam, the type of questions he wants to ask (i.e. selections only or ordering only or a combination of selections and ordering).

Additionally, sometimes the assessor would like to give more weight to the selection part than to the logical order part or vice versa. In this case, the assessor should determine which part he would like to weight, while he is writing the exam, in the space provided in Create Exam Component. The spaces are called Selective Multiplier (for the selection part) and Logical Multiplier (for logical order part). For example, suppose that the assessor would like to give more weight to the selection part, then he should type a larger number in the 'Selective Multiplier' (e.g. 3) and a smaller number in the 'Logical Multiplier' (e.g. 1). Nevertheless, if the assessor would like to give the same weight for both parts, then he should type the same number (e.g. 1) in both multipliers.

ii. Sit Exam Component:

Students can practise the exam from this component (Figure 3.6, page 128). Initially, the student should enter his name so that the program can keep a record and store the results of his answers. Then, an introductory paragraph(s) will appear instructing the student how to answer the exam.

The following steps show how the student can answer the exam:

- Student should select the number of the question he wants to answer (Q1, Q2, Q3,....etc.) from the top bar.
- Double click on the options he would like to include, in his answer, will go automatically to the place where the answers should be put.
- If the student changes his mind or makes a mistakes, there is an 'UNDO' facility which enables him to return and make the changes.
- If the question asks the student to order his selection logically, then the student should select the 'ORDER' icon after he makes the selection. This icon enables him to shuffle the options to reach the best order.

- If the student wants to finish the exam he should click on 'FINISH' icon. After he does so, the program will ask him a question to make sure that student really wants to end the exam.

iii. Edit Exam Component;

This component allows the assessor to edit an exam already in the program. Sometimes, the assessor, for some reason, would like to add or remove some questions from the existing exam. He can do this very easily by clicking on this component and do the editing. After all the editing is done the assessor should save the exam.

3.3.5.3 Sample to Test the Program:

Designing the program is not the end of the story. It is necessary to give it a trial. Therefore, the program was supplied to four high schools in Scotland to try with some test materials in biology and chemistry for S4 pupils as a tool for self – assessment (Appendix 5.2).

Three months later, a letter was sent to every principal teacher of biology and chemistry who received the CD and the testing materials of the four high schools. The purpose of this letter was to know how they found the program in terms of installation, the usage, the layout, the weakness and any comments or suggestions for improvements.

Unfortunately, the researcher did not receive any reply from any teacher. It may be that they do not have time to try the program or something related to the school policy.

3.3.5.4 Limitations of the Computer Program for SCGs

The computer program for SCGs is a good starting point to deliver the grid questions in an accessible way for teachers and students. It is our intention to see this program widely used by many schools and universities.

However, the program has some limitations and is in need of some improvements. This is because it is a new program and each new program needs an evolution and improvement until it appears in the way that it can be used efficiently.

Some limitations of the program from the researcher's point of view are as follows:

- The program can only be used in a PC computer. Some institutions prefer to use Macintosh computers in their departments. Thus, these institutions will not be able to use the program.
- The assessor can design as many questions as he can based on one grid but he should use only one file for each grid. For example, suppose that the assessor would like to set up an exam consisting of four grids he should save it in four files (i.e. one file for each grid). In order for pupils to answer the exam, they should open the four files, which may in some cases affect pupils' responses.
- The assessor cannot design questions based on formulae, equations and shapes which might be desirable in chemistry and physics. The program does not allow the assessor to write options having chemical formula, chemical equations and molecular geometry in chemistry, for example. The only solution for this is to convert these things into words, if possible, otherwise the assessor should avoid asking questions like this.

Example of the Question which cannot be asked in the Program:

Ion-electron equations can be used to show the gain and loss of electrons in chemical reactions.

$Fe_{(s)} \rightarrow Fe^{2+}_{(aq)} + 2e$ **A**	$Fe^{2+}_{(aq)} + 2e \rightarrow Fe_{(s)}$ **B**	$Fe^{2+}_{(aq)} \rightarrow Fe^{3+}_{(aq)} + e$ **C**
$Fe^{3+}_{(aq)} + e \rightarrow Fe^{2+}_{(aq)}$ **D**	$Cu_{(s)} \rightarrow Cu^{2+}_{(aq)} + 2e$ **E**	$Cu^{2+}_{(aq)} + 2e \rightarrow Cu_{(s)}$ **F**

a. Select the box contains the equation which shows iron (II) ions being oxidised ?

____

b. Select the box(es) containing the two equations which show the reactions which occur when an iron nail is placed in copper (II) sulphate solutions?

____

- There is only a limited number of boxes in the program (maximum 12 boxes). The assessor cannot exceed this number and ask questions based on a grid with 14 boxes, for example: At the university level, some researches such as Johnstone (1988) successfully used grid questions with sixteen boxes. Therefore, grid questions with sixteen boxes cannot be used to assess fourth year students by the program.

These are some limitations of the program from the researcher's point of view. Nonetheless, an improvement will be necessary to overcome these limitations so that the institutions can benefit from the facility offered by the program to improve their assessment process.

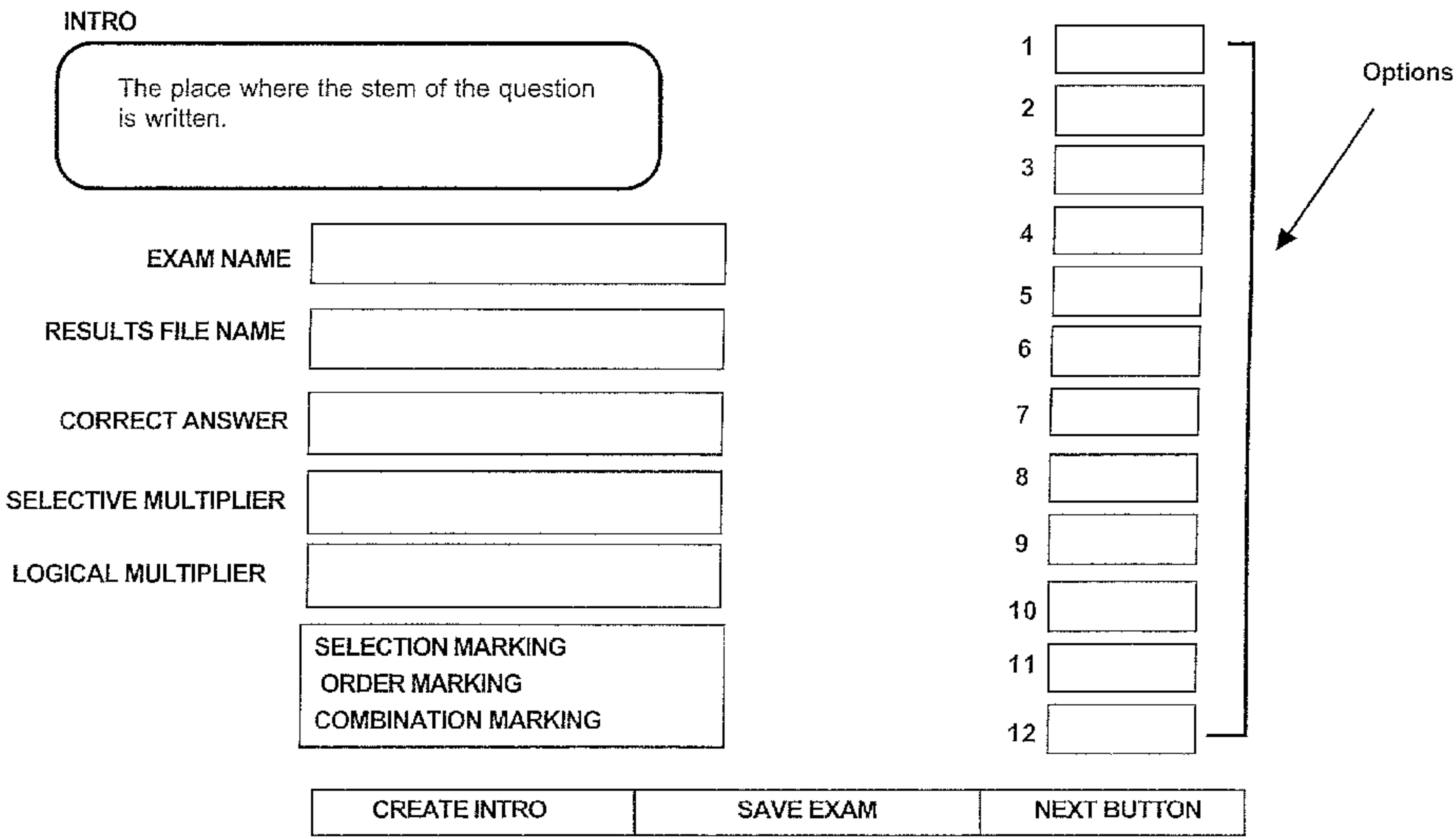


Figure 3.5: The layout of the Create Exam component of the SCGs computer program

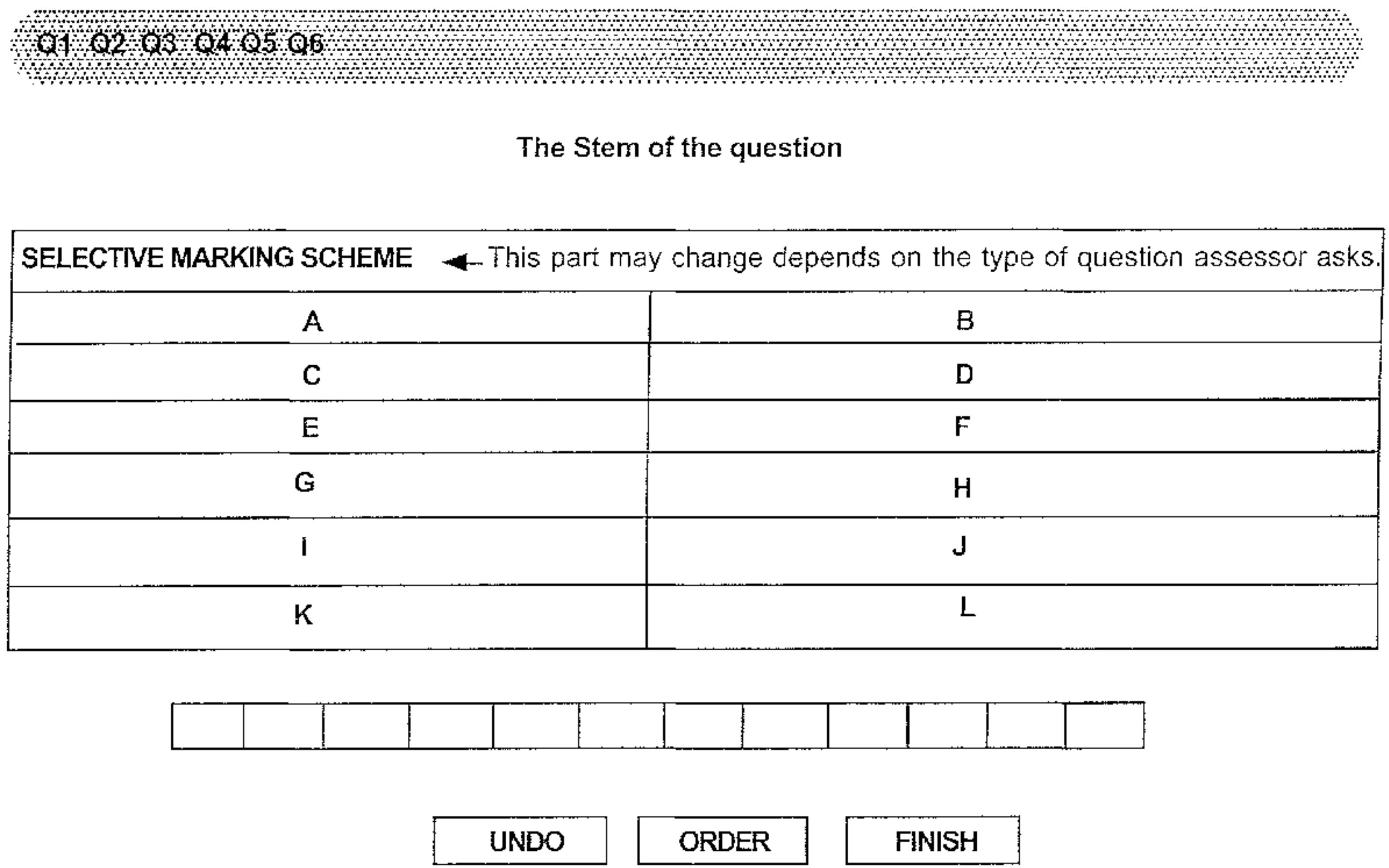


Figure 3.6: The layout of the Sit Exam component of SCGs computer program

CHAPTER FOUR: LITERATURE REVIEW (2)

CHAPTER FOUR: LITERATURE REVIEW (2)

4.1 Cognitive Styles and Assessment Methods:

At the beginning of this section we need to ask ourselves the following **question**; "*Are there any type(s) of assessment methods which one group of cognitive styles might perform better in than others?*" As a background of this part of the current research, the researcher is going to review what has been written in the literature about cognitive styles and its relationship to assessment methods.

4.1.1 Introduction

Cognitive styles (also called Learning or thinking styles) have a great influence in the learning and teaching process. They can be defined as relatively stable, self - consistent modes of adaptation that mediate the ways in which individuals process information (Ausubel, 1963; Brodzinsky, 1982). There are different cognitive styles such as field - dependence versus field - independence, convergence versus divergence, complexity versus simplicity, analytical versus global. This research is attempting to focus on the first two styles; field dependence - independence and convergence - divergence because they have an important role in the educational process.

There is a wealth of research in the area of the relationship between cognitive styles and the learning and teaching process. However, there is a lack of research in the area of the relationship between cognitive styles and the assessment process. It is one aim of this research to investigate such a relationship.

4.1.2 Field Dependent - Independent Cognitive Style:

4.1.2.1 Description of Field Dependence - Independence:

For a long time field dependent - independent cognitive style has been seen as an important factor in the educational process. It is the most investigated area of the cognitive styles. This is due to the fact that field dependent - independent cognitive style is appeared in the majority of psychological domains, providing a coherent and consistent reflection of individual behaviour differences (Tinajero *et al.*, 1993). Furthermore, it is related to overall academic achievement (Tinajero *et al.*, 1998). Much research has been done which tried to investigate the effect of field dependent - independent cognitive style on students' learning.

The early work in the area of field dependence - independence was carried out by Witkin and his colleagues (Witkin *et al.*, 1977; Witkin *et al.*, 1981). Then, further research has been done to reveal and examine more related points.

Field - independent learners (FIs) can be described as those who can easily break up an organised perceptual field and separate an item of information from its background (Witkin *et al.*, 1981). In simple words, field - independent people are those who can separate **'signals'** from **'noise'** or separate **'relevant information'** from **'irrelevant information'**. On the other hand, a field - dependent learner is one who has difficulty in breaking up an organised field and separating an item of information from its background.

4.1.2.2 The Differences between Field - Dependence and Field - Independence:

Generally speaking, field - dependent learners (FDs) seem to have more difficulty than field - independent learners (FIs) in determining the structure and separating the relevant information in a given complex array of information (Rickards *et al.*, 1997). It is due to the fact that a field - independent person is more capable of cognitive restructuring skills than a field - dependent person. This fact is supported by the work of Noble *et al.* (1985) who found that a field - independent person is better at cognitive restructuring tasks.

The cognitive restructuring may involve three separate but related operations (Witkin *et al.*, 1977; Al - Naeme, 1991).

1. The ability to break up the task into its task elements.
2. The ability to manufacture a structure from an ambiguous stimulus complex which will be the outcome of such procedures.
3. The ability to make a different organisation of the task than its initial structure in the stimulus complex.

Furthermore, field - independent individuals show more interest in analytical - impersonal domains and restructuring. Field - dependent individuals, on the other hand, show an interest in global - interpersonal fields in particular those which require social skills (Alamolhodaei, 1996; Tinajero *et al.*, 1997). Moreover, Tinajero *et al.* (1997) state that field - independent learners 'trust' internal cues and this is associated with greater aptitude for restructuring i.e. for imposing organisation on received information. In contrast, field - dependent learners place their 'trust' in external cues, and tend to accept percepts, or symbolic representations at face value. Also it is worth noting that field - dependent learners tend to accept the organisation of the material to be learned as given, while field - independent learners tend to actively manipulate the material to be learned (Davis *et al.*, 1979).

4.1.2.3 Field Dependence - Independence and Assessment in Science Education:

In the assessment process, it seems that this field of cognitive styles presents a problem to the examiner. The examiner may face a problem when he asks open - ended questions for those learners who have experienced difficulty in structuring (Field - Dependent). On the other hand, he faces a problem when he asks specific questions with those learners who can do their own structuring (field - independent). However, using various assessment methods might help to solve the problem.

The other related point is that, within the context of the multiple choice test, perhaps field - independent examinees would be more effective when they ask to change their answers than those who are field - dependent. This is due to the fact that field - independent persons tend to be more analytical (Friedman *et al.*, 1995).

The results of Hasson's study (1988) asserted that low capacity field - dependent students seemed to do better in fixed response questions than in open questions. Low capacity students are those who have a capacity score of 4 or 5 in the Digit Span Tests; the tests which are used to measure human working memory capacity (Hasson, 1988; Johnstone *et al.*, 1989).

4.1.3 Convergent - Divergent Cognitive Style:

4.1.3.1 Description of Convergent - Divergent Cognitive Style:

The early appearance of the convergent - divergent cognitive style was from the work of Hudson. After reviewing the work of Getzels and Jackson and his own work, Hudson (1962) called the high **IQ** people as convergers and the highly creative as divergers. The term 'highly creative' was replaced by the words 'open ended test' later by Hudson.

Hudson (1962) classified his sample, after analysing the data, into three main categories (Figure 4.1).

1. Convergers: are those who are substantially better at an intelligence test (**IQ**) than they are at an open – ended test. He found that 30% of his sample were convergers.
2. Divergers: are those who are substantially better at an open – ended test than they

are at the intelligence test (**IQ**). He found that 30% of his sample were divergers.

3. All – Rounders: are those who are more or less equally good (or bad) on both types of the test. He found that 40% of his sample were all – rounders.

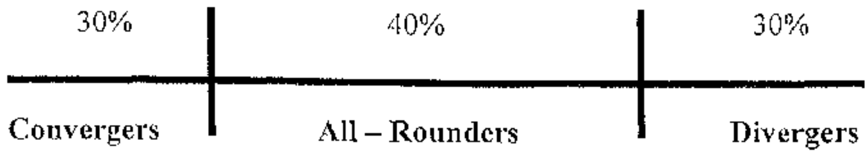


Figure 4.1: The three main categories of Hudson's classification of the convergent - divergent cognitive style

4.1.3.2 Summary of Hudson's Work:

The following points summarise Hudson's main findings:

1. The convergence - divergence dimension is a measure of bias, not of level of ability. There was a belief among some psychologists that divergent people were potentially creative and convergent people were potentially unreactive. However, this view was refuted by Hudson who saw that convergers are naturally attracted towards one end of the spectrum and divergers to the other.

2. Most arts specialists were weak in **IQ** tests, better at open – ended tests (Divergers). On the other hand, most science specialists were weak at open – ended tests, better at **IQ** test (Convergers). This is may be due to the fact that science does not consist so much in the free play of the imagination, even over the practical matters. The study by Lloyd - Bostock (1979) suggested that there is a clear relationship between arts orientation and divergent thinking before any specialisation has taken place. Runco (1986) supported Lloyd - Bostock's (1979) study and states that there were particular areas of performance (e.g. writing and art) that were more strongly related to divergent thinking than other areas (e.g. music and science).

Nevertheless, does divergent thinking ability discriminate between specialisation in arts and science? Webster *et al.* (1981) recommended that divergent thinking ability, which is used to discriminate between specialisation in arts and science, must be treated with caution. They found in their study that item content has a significant effect on divergent production. For examples, the studies by Al - Naeme (1991) and Alamolhodaei (1996) showed that when divergent thinkers take science or maths they perform better than convergent thinkers. It seems that the divergent thinkers are more capable than others in finding the way to an answer. In addition, they are more skilful in giving a variety of answers to every day life questions (Runco, 1986; Al - Naeme, 1991). The convergent thinkers, by contrast, have difficulty in giving a variety of answers to every day life questions.

4.1.3.3 Convergent - Divergent Cognitive Style and Assessment in Science Education:

Unfortunately, little research has been done in the area of assessment. However, one of the disadvantages of MCQs and most of the fixed response questions is that they do not give the opportunity to the divergent thinkers to speculate and perhaps give an answer not considered by the examiner (Jackson, 1988). In MCQs students are presented with four or five options only in the item. On the other hand, divergent thinkers are expected to perform better than convergent thinkers in open response questions. This fact is supported from the previous studies which concluded that divergent thinkers are better at open - ended questions. Furthermore, in open response questions students have the opportunity to express their ideas and show their ability.

The divergent thinker examinees are looking for more possibilities. This can be found in some fixed response methods of assessment such as structural communication grids (SCGs) (for more information about SCGs see Chapter Two, section 2.2.1.5). In SCGs students are presented with more options than in MCQs in one question. Therefore, it could be hypothesised that divergent thinkers would perform better than convergent thinkers in SCGs. The divergent thinkers may have better chance to show their ability within the options given in SCGs.

Bahar (1999) examined students' performance on SCGs and their performance in convergent - divergent thinking style test. He found that students who had divergent thinking style had higher scores than the students who had a convergent thinking style. One aim of the current research is trying to find out if the same results that Bahar's came about can be found.

4.2 The Representation of Knowledge in the Pupil's Mind:

4.2.1 Introduction:

After reviewing, in brief, what is written in the literature about some cognitive styles and their relationship to assessment methods, the time now is to review, also in brief, what is written in the literature about pupils' knowledge structures. The **key question** in this part is that "*Is there any relationship between the way in which pupil's knowledge is stored and the test structure based on MCQs?*".

This section is attempting to present some theoretical background for how knowledge is represented in the pupils' minds. This might help to explain the results which have been found in some parts of the current research. The key point which has been suggested in all the associative theories is that there is a connection between the materials which have to be learned and the things one already knows. These connections provide retrieval paths, routes that one can take when searching for some particular information in memory (Reisberg, 1997). They provide the basic building blocks through which knowledge is represented in the memory.

There are two types of memory in the human mind; Semantic memory and Episodic memory (Tulving *et al.*, 1982). Ashcraft (1994) describes semantic memory as our permanent memory store of general world knowledge, variously described as a thesaurus, a dictionary, or an encyclopedia (i.e. where knowledge is stored). Whereas, episodic memory is a personal, autographical store (where experiences are stored). For example, one person's memories of a university are quite different from another person but they share a general kind of knowledge about being a university student. The discussion of the following sections will be concentrated on the semantic memory because it might be expected to be more directly relevant to the practice of science teaching (Preece, 1978).

4.2.2 The Cognitive Structures:

It is important that the term **'cognitive structures'** is clearly defined and understandable by the reader. Shavelson (1972) describes the word **'structures'** as an assemblage of identifiable elements and the relationships between those elements. When talking about the structures of these elements in the teacher's or pupil's memory then the term **'cognitive structure'** has been used. Therefore, cognitive structures refer to the relationships or organizations of the concepts in the memory. There is a need to map the cognitive structure of pupils' minds of certain scientific concepts to obtain structural information about the internal representation of the concepts.

Most of the work on associations comes from the learning point of view. During the processes of teaching and learning, a pupil learns more concepts. These concepts are linked to the existing ones to form a network in the pupil's mind. The new and existing concepts now become meaningful (Johnson, 1964; Shavelson, 1973). Then, the pupil rearranges them in such away that they can be stored in memory more efficiently. One way of storing these concepts is to put them in an equation like in physics or mathematics. In this case, a large amount of information is stored in memory as a small amount. However, if the pupil has not acquired the meaning of the concepts, then he faces difficulty in finding an efficient method for chunking. Chunking is a process of organizing information which allows a number of items to be viewed as a single unit; e.g. the digits 1 9 8 7 could be viewed as four units or as a date, 1987, which is a single unit of information (Cassells, 1991).

The meaning of any word cannot be determined unless it is looked at from its relations with other words and other objects in the natural word (Deese, 1962; Johnson, 1967; Johnson, 1969). Shavelson (1973) argues that meaningfulness is directly proportional to the number of associations given to a stimulus word. In order to make a judgement about the association of say word **X** to word **Y**, then one should look to all the responses that the two stimulus words have in common. If the two stimuli have identical responses, then it can be assumed that the two stimuli are strongly associated with each other.

4.2.3 Models of Semantic Memory:

There are different models which try to describe how knowledge is represented in the human memory. Examples are the Hierarchical Network Model, Feature List Model and the Connectionism Model. The discussion will cover only the first model (Hierarchical Network Model) as an example of these models because it is simple and meets with the way this research was designed.

4.2.3.1 Collins and Qullian (and Loftus) Model (Hierarchical Network Model):

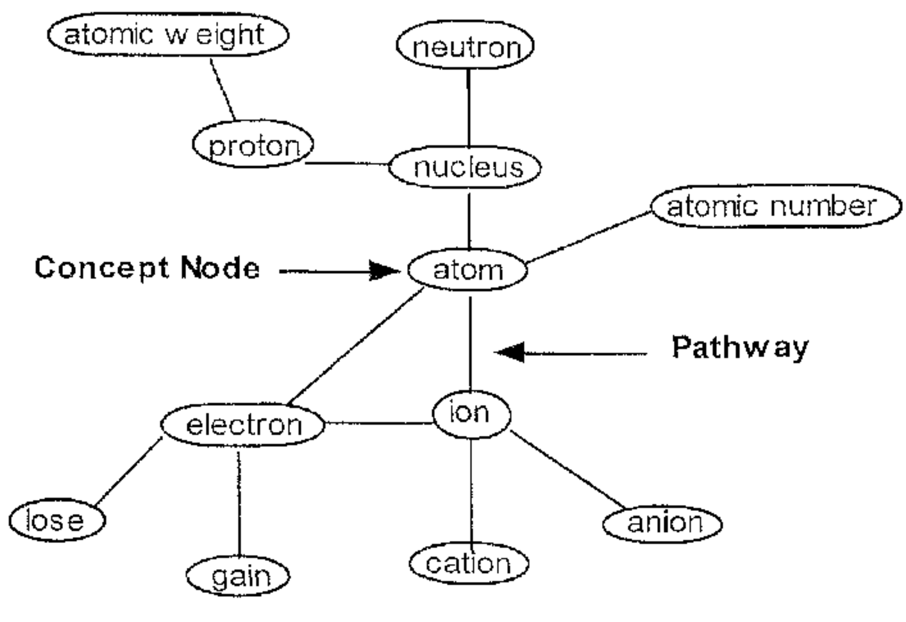
This is the first model in cognitive psychology which tries to explain the structure and processes of semantic memory (Ashcraft, 1994). There are two assumptions which could be made in this model. The first assumption is about the structure of semantic memory and the second assumption is about the process of retrieving the information from that structure. In Collins and Qullian (and Loftus) Model, the learner should play a role. That is, the connections between the concepts will be established only if the learner pays attention to what is to be remembered during the learning episode (Reisberg, 1997).

i. The Structure of the Semantic Memory:

Collins and Qullian (1969) suggested that semantic memory is a network or interrelated set of concepts, or interrelated body of knowledge. Their view about the semantic memory was driven from an earlier computer simulation model. Later, another theory emerged from the work of Collins and Qullian. It is a theory of spreading - activation which could be used to explain human performance (Chang, 1986).

In the Collins and Quillain (and Loftus) model, the knowledge is stored in memory in the form of an associative (semantic) network, with the nodes in this network corresponding to word concepts (e.g. bird, feathers) and the links to conceptual relations joining these concepts (e.g. ***is - a*** for the subset - superset relations; ***has, can,*** or ***is*** for attribute relations). Each concept node is linked to another concept node by a pathway (Figure 4.2). The pathway, however, is not necessarily short and direct. Indirect and long pathways can eventually be traced between any two nodes. The full meaning of any concept is the whole network as entered from the concept node (Collins *et al.*, 1975)

Figure 4.2: A simple example of Collins and Qullian (and Loftus) Model of semantic memory from chemistry



ii. The Process of Retrieving the Information:

After knowing the structure of the semantic memory, how does the process of retrieving the required information take place? The process of operating in the semantic memory that leads to retrieving the information is called **spreading - activation**. Spreading - activation is the mental activity of accessing and retrieving information from the network. It is the process by which activation spreads from node to node along network link, making knowledge associated with particular sources of activation available for processing (Anderson *et al.*, 1984).

One of the most important aspects of the activation process is that it spreads through the network. The activation begins with one concept node that represents what the sentence is asking and then spreads throughout the network along all the connection pathways until the target information is reached (Reisberg, 1997).

The activation level of each node depends on two factors; how much activation the node has received and how recently the activation occurred. In this model, the word activate sometimes is replaced by the word *prime*. However, what will happen to the activation after it has activated certain nodes? It was found by experiment and data analyses that activation tends to decay exponentially with distance from source (Anderson *et al.*, 1984).

What about if two concepts become activated. In this case, there are two simultaneous spreads of activation one from each concept node. A connecting route or set of pathways has been retrieved from the semantic memory. The spreads of activation will encounter one another. When this happen, an intersection has been found between the two concepts. After the intersection has occurred, a decision stage must operate to make sure that the retrieved pathway is valid. What is meant by valid in this case is that it represents the relationship specified in the sentence. The benefit of spreading two simultaneous activations is helping to raise the target node to threshold levels. An example is if a student is given a hint or clue in the sentence. A study by Bowden (1985) suggested that informing subjects of the relation between relevant information and a problem's solution (clue) facilitates problem solving by allowing them to access the information more rapidly. It could be explained that giving a subject a clue sentence will direct the searching process and thereby decrease the time necessary to access relevant information.

However, spreading activation is not only trying to retrieve the relevant pathway between the two concepts but also activates all the related concepts. These related concepts will not be activated forever, but this activation will decay after some time.

4.2.4 The Empirical Tests of Semantic Memory Models:

It is time to examine the test findings of the semantic memory models. The early tests of semantic memory were made by what is called a *sentence verification task*, in which simple sentences are presented for the subjects' yes/no decisions (Ashcraft, 1994). These tests rely on the assumption that people have the relevant knowledge in memory. In addition, such semantic knowledge is largely similar among individuals or individuals within a language culture. The following points list the results of these tests:

a. Collins and Quillian (1969) concluded that the two concepts that are closer together in the network should require less time for verification than the two that are farther apart. Therefore, the reaction time depends on the distance between the concepts in the semantic memory. As the distance between the two concepts increases, there will be an increase in the reaction time (Figure 4.3).

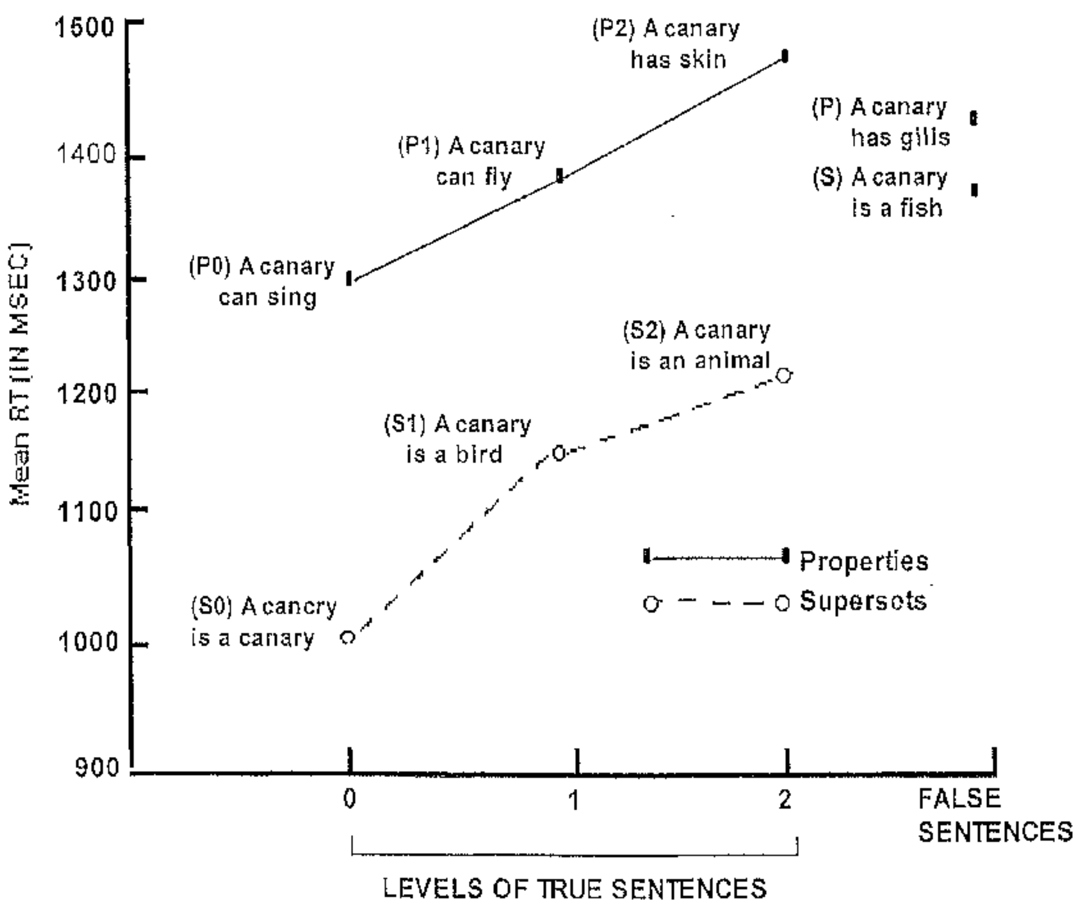


Figure 4.3: Reaction time (RT) to superordinate (S) sentences and property (P) sentences is shown as a functional of levels within the hierarchy. An S2 sentence involves a superordinate connection two levels up the hierarchy; S1 means one level up in the hierarchy; a 0 level sentence had the predicate stored at the same hierarchical level (After Collins and Quillian, 1969)

In the above figure, Collins and Quillian (1969) used the symbol (S) to indicate a superordinate statement and a (P) symbol to indicate the property statements. The values of the **X**-axis indicate how many levels in the hierarchy the search had to proceeds through in order to find the stated concepts.

From the figure, the reaction time increased as the semantic distance between the two concepts increased. For superordinate sentences, searching for S2 relationship (e.g. *A canary is an animal*) required longer time than searching for an S1 relationship (e.g. *A canary is a bird*). The same thing (i.e. increase in time) was found for the property sentences going up 1 versus 2 levels.

Moreover, it appeared from Collins and Quillian results that the *isa* pathways were stronger than property (attribute) pathways since superordinate sentences were faster overall than property sentences (Aschcraft, 1994). The final conclusion, which could be summarised from above, that semantic structure is based on the semantic relatedness among concepts.

b. Memory Cognitive Economy: The model which was proposed by Collins and Quillian (1969) assumed properties, which do not uniquely define a word but are also properties of the word's superordinate, are stored only in the configuration which defines the meaning of the superordinate (Conrad, 1972). In other words, the properties need not be stored at all nodes to which they apply, but just at their most general node, as this automatically entails that these properties also belong to concepts at a more basic level (Chang, 1986). For example, the property 'flies' would not be stored with *canary*, but rather with *bird*. Therefore, to make a connection, if the *canary* can fly, comes from the fact that *canary* is a *bird* and *birds* can fly. Another example is the property 'eats'. This property only needs to be stored with animal, because the fact that *bird*(s) eat or *fish* eat(s) can be derived from the fact that *bird* and *fish* are subdivision of *animal*. This storage in hierarchies makes the memory more economical. To make the memory most economical in the number of concepts which must be stored in the semantic memory, only nonredundant facts will be stored in memory.

However, some psychologists do not support this view. A study by Conrad (1972) concluded that properties are stored in memory with every word which they define and can be retrieved directly rather than through a process of inference. It is possible that Collins and Quillian had obtained their results not because the properties or superordinates were more distant in a hierarchical structure, but instead because their "more distant" concepts were actually associated more weakly with the concepts (Aschcraft, 1994).

c. Typicality Effects: Sometimes a subject is responding to one sentence more than another, although the connection between the two things in the semantic memory is the same. This can be explained by what is called the typicality of what to remember. For example, the *robin* is generally considered more representative of '*bird*' category than is a *chicken* and a statement ***'a robin is a bird'*** is verified faster than statement ***'a chicken is a bird'***. A study by Batting *et al.* (1969) suggested that pathways to less common category members were longer, making those members further away in the semantic network. If there are typical members and atypical members in the category, the typical members can be judged faster than atypical ones. The study by Collins *et al.* (1975) supported the Batting *et al.* (1969) study and concluded that the typicality effect is one more manifestation of the fact that semantic similarity speeds up positive decisions and slows down negative decisions.

d. Degree of Fan: This is one factor which affects the retrieval process in the semantic memory. It can be defined as how many things are fanning out from the subject (Figure 4.4). The subject knows only one fact about something (low degree of fan). This fact will be easily remembered because all the activation from the concept node will spread down to that single fact. But what about if the subject knows many facts about something (high degree of fan). One would think that there is a disadvantage in this because the activation will spread to different pathways and hence slow down the response time. It could be true, but knowing many things also has an advantage. The advantage of knowing many things is that there exist multiple retrieval paths for the information and consequently this aids the retrieval process when the learner learns new facts about something.

Degree of Fan for Two Nodes

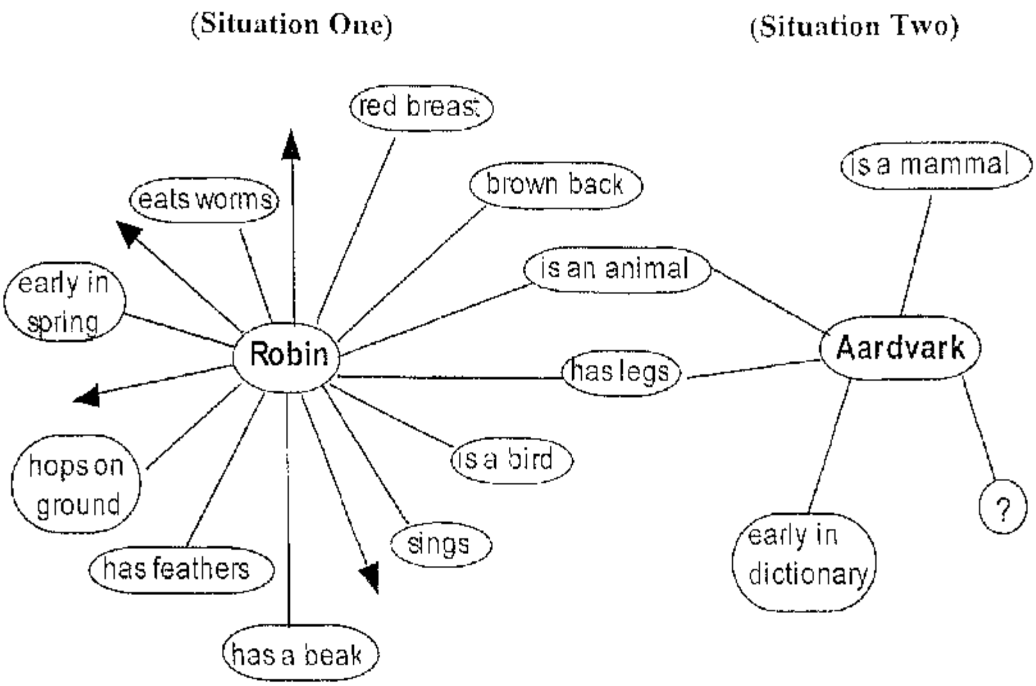


Figure 4.4: Shows the two situations. Situation one represents high degree of fan (many linkages radiating out from the ROBIN node). Whereas, situation two represents low degree of fan (few linkages radiating out from the AARDVARK node) (After Reisberg, 1997)

4.2.5 Retrieval Blocks and Partial Retrievals:

Sometimes a subject has difficulty or is unable to come up with a certain word or a sentence that he surely knows, or may retrieve only part of the required information. The former case is called a retrieval block whereas the latter is called partial retrieval. Retrieval blocks refer to experiences in which target information available in the memory store is temporarily blocked from retrieval by the retrieval of closely related but erroneous information (Roediger *et al.*, 1982). For example, one who tries to remember or recall the name of an animal (name of a dog or a cat). The desired name does not come to mind, but another name comes to mind during repeated retrieval attempts. The erroneously retrieved name has the same initial letter as the correct target and shares some properties.

Retrieval blocks are one of the problems that several associative theories of semantic memory have to face. Although there is no clear explanation about how the retrieval blocking occurs, there are some ideas of how the blockage happens. One explanation suggests that the target nodes are receiving a great deal of activation, given the time and effort one spends in trying to find the word. But as Reisberg (1997) argues that this activation seems not to spread to the sought - after node or, if it does, the activation of this node is not being recognized or acknowledged. Under certain conditions retrieval becomes hampered rather than aided by the presence of information that is related to the target information.

4.2.6 Cognitive Structures and the Assessment Process:

Little attention has been given to why some pupils fail to answer examination questions from the cognitive point of view. Pressures on teachers can sometimes lead to an emphasis on how many pupils pass their courses, neglecting to consider why some pupils fail.

Kempa *et al.* (1983) assert that the inability of the pupils to solve problems is not related or linked to pupils' general intellectual skills or their use of adequate or inadequate problem - solving strategies. It is related, however, to the cognitive structures of scientific concepts in pupils' minds. Kempa *et al.* (1976) made two studies on why some pupils failed the course and reached the same results. The results of the first study hypothesised that pupils' ability or failure to cope with chemical problems depends on their level of concept attainment and the association between the scientific concepts. The results of Kempa *et al.* (1983) second study showed that where the concepts are only weakly linked, access from one concept to another is not readily achieved. Thus, the problem - solving which requires the link does not occur.

4.3 Word Association: A Technique of Revealing the Cognitive Structures:

The reason for writing a section in the literature part on the word association technique is because this technique was used in the field research as a tool for revealing the pupils' cognitive structures. In addition, it might be that the new reader is or is not that much familiar with this technique.

4.3.1 Description of the Word Association Technique (WAT):

After knowing how the concepts can be represented in pupils' minds, the following questions are looking for answers.

a. How accessible are certain concepts in the pupils' cognitive structures?

b. How can the teacher gain a useful insight into pupils' present ideas?

c. How can the teacher identify the cognitive structures of his pupils?

The answers to the above questions will be discussed in the following paragraphs. It was argued that cognitive structure is not directly accessible through experimentation and observation (Kempa *et al.*, 1983). Nonetheless, there are some methods which can be used to investigate such structures. Most of these methods have a problem of validity and reliability and also they are often very time - consuming (Cachapuz *et al.*, 1987). Examples of these methods are word association, structural communication grids, tree construction and concept mapping.

Word association technique (WAT) is one of the oldest methods for identifying cognitive structures. It is a direct probe of the associations that a person perceives for a set of concepts (White *et al.*, 1992). In this technique, pupils are asked to give a series of one word responses in response to a given term (called the stimulus) in a limited time. For example, pupils were asked to give one word responses (ten words) to a stimulus word '**Gas**' in chemistry. The following are the responses given by pupil **X**:

Example of WAT from Chemistry:

GAS

GAS	1. Hydrogen
GAS	2. Air
GAS	3. Element
GAS	4. Oxygen
GAS	5. Pressure
GAS	6. Temperature
GAS	7. Carbon Dioxide
GAS	8. Reaction
GAS	9. Compound
GAS	10. Car

The underlying assumption of word association technique is that the order of response retrieval from long - term memory reflects at least a significant part of the structure within and between concepts (Shavelson, 1972). As a result of this, the order of responses generated by each pupil to each concept takes on particular importance.

There are two types of word association; controlled and free word association (Preece, 1976). In controlled word association, the responses are restricted to a given category whereas in a free test no such restriction is imposed. Preece (1978) argues that restricting associations to the semantic domain involved was a valuable technique in mapping cognitive structure.

4.3.2 Advantages of the Word Association Technique:

The word association technique is one of the most popular among researchers trying to map pupils' cognitive structures. This is because of many advantages such as follows:

a. It can be a useful tool for revealing the type and the number of concepts in the pupils' minds along with the links existing between them (Johnstone *et al.*, 1985; Bahar *et al.*, 1999).

b. It is simple to prepare and administer. WAT does not take a long time to apply and can be used for large number of pupils. It does not require complicated instructions to apply and hence the class teacher can use it.

c. It was felt to be the most objective method of preparing a suitable achievement test (Johnstone *et al.*, 1985). Teachers can use the responses which were given by the pupils in the WAT as distractors in the MCQs.

d. It can be used to compare a pupil's understanding with a standard such as the connections that a textbook specifies between concepts i.e. Content Structure (White *et al.*, 1992).

e. It can be used to compare the understandings of an individual pupil or a group of pupils to the given concepts of the topic.

f. It allows an operational definition of connectedness, and is amenable to quantification. This advantage is the one that attracts many teachers and researchers to use WAT. Calculating relatedness coefficient, for example, can be used to determine the overlap lists triggered by different stimulus words (Sutton, 1980).

4.3.3 Disadvantages of the Word Association Technique:

The word association technique has some limitations for use as a tool for investigating the cognitive structures of the pupils' minds.

a. No decision can be made in the interpretation of cognitive structures as to the connectness or otherwise of association since the pupil's reasons for making the association are not known (Kempa *et al.*, 1983). Even assigning numerical values for pupil's responses will not eliminate the subjectivity of the interpretation of this technique.

b. A pupil may know the response words for the given stimulus word but there is no guarantee that the pupil understands these responses. It is like when pupils may know their way round the library without understanding the books.

c. Teachers may get different types of response to the given stimulus word (Sutton, 1980). For example, nouns and adjectives, words that are associated just because they sound similar, or paired opposites, words that are similar in meaning or ones that are used together, but the pupil does not know why (for example electron - proton, acid - alkali, vertebrate - invertebrate). To minimize the effect of this problem, the teacher may limit the situation by some instruction about acceptable responses in terms of bonding theory in chemistry, physical quantities in physics and reptiles in biology.

d. The semantic distance which is calculated in WAT only tells the teacher that one idea does trigger another but it does not tell what is the relationship between the ideas.

4.3.4 Scoring Word Association Tests:

There are three properties that are of particular importance in interpreting the word association data (Shavelson, 1974). These are the number of responses to each stimulus word, the nature of these responses and the overlap between responses to pairs of words. They can be converted to scores so that each pupil can be allocated a raw score in the test.

4.3.4.1 Number of Responses:

This property indicates if the pupil understands the given stimulus word to some extent. It could be assumed that the greater the number of responses the better the understanding. The reason behind this assumption is that while the pupil learns the subjects, the key concepts should increase in meaningfulness. Consequently, the average number of responses to each concept should increase.

However, one must be cautious about using the number of responses as a measure of understanding the given key words. The layout of the response sheet may constrain the number and that number is a function of the person's fluency (White *et al.*, 1992). For example, suppose that there are two pupils; one gave fewer responses than the other. This does not mean that the pupil who gave more responses has a better understanding than the pupil who gave fewer responses.

Despite the above reservation, if the teacher is willing to use the number of responses as a measure of understanding, then he may allocate one mark for each given response which is relevant to the piece of learning.

4.3.4.2 Type of Responses:

The teacher has to have in mind some categories of association or a set of criteria for the type of responses to be counted. In this case, the teacher may allocate one mark if the pupil gave a response word which is judged as acceptable for the topic that stimuli words have driven (Johnstone *et al.*, 1985). There are two views about this method. One view sees the pupil who gave words which are outside the topic as a creative pupil who should not be penalised. The other view, which is opposite, sees the outside associations as a reflection of disorganised knowledge or random listing and should be penalised. It is up to the teacher to decide which view he takes and which types of response indicate good understanding.

4.3.4.3 The Overlap in Responses Lists:

This method is looking to the WAT data from the quantitative point of view (Shavelson, 1974). In this property, two things are considered when scoring pupil's responses. These are the common responses between the two stimulus words and the rank order of responses in the two response lists. The order of the associative words obtained has an important point in determining the meaning of the word. The first responses carry a grater share of meaning than later ones (Garskoff *et al.*, 1963). It could be assumed that calculating the degree of overlap of response hierarchies is a measure of the semantic proximity of the stimulus words (Bahar *et al.*, 1999).

One of the oldest and commonest methods of calculating the overlapping is the "Relatedness Coefficient (RC)" developed by Garskoff and Houston (1963). The Relatedness Coefficient is measuring the overlap and the correspondence in order. The following formula is a simplified equation of Garskoff and Houston (White *et al.*, 1992).

$$\text{Relatedness Coefficient (RC)} = \frac{\text{Sum of products of ranks of common words}}{(\sum n^2) - 1}$$

Where n is the number of words in the longer list. The denominator would signify perfect relatedness.

The value of the relatedness coefficient ranges between 0 and 1. If the two lists are identical in both the common responses and in the rank order, then it is assumed that the two stimulus words are strongly associated (RC = 1, perfect relatedness). On the other hand, if the value of relatedness coefficient is equal to zero, then it could be assumed that there are no links or associations between the two stimulus words (totally unrelated). The sequence of the common words on the two lists affects the value of relatedness coefficient.

To make the formula easy to understand, an example will be given. Suppose that pupils are given different stimulus words on the cell division topic in biology. "Mitosis" and "Meiosis" are the two stimulus words in that topic. One pupil gave responses to the two stimulus words as shown in Table 4.1.

Table 4.1: The responses given by one pupil of Mitosis and Meiosis as two stimulus words in Cell Division topic

Associates to MITOSIS	RANK	Associates to MEIOSIS	RANK
Mitosis	9	Meiosis	8
Zygote	8	Chromosome	7
DNA	7	diploid	6
Nucleus	6	Zygote	5
Chromosome	5	haploid	4
Meiosis	4	Chromatid	3
Chromatid	3	Mitosis	2
Cytoplasm	2	DNA	1
Copying	1	-	-

The common response words are mitosis (9,2), zygote (8,5), DNA (7,1), chromosome (5,7) meiosis (4,8) and chromatid (3,3). From the above table, mitosis has the longer list.

$$RC = \frac{(9*2) + (8*5) + (7*1) + (5*7) + (4*8) + (3*3)}{(9^2 + 8^2 + 7^2 + 6^2 + 5^2 + 4^2 + 3^2 + 2^2 + 1^2) - 1} = \frac{141}{284} = 0.50$$

After calculating the relatedness coefficient for each pupil for each pair of stimulus words, the teacher may calculate the relatedness coefficient for the whole class. This can be done very easily by calculating the mean relatedness coefficient for each pair of the stimulus word. Calculating the mean relatedness coefficient helps the teacher to map the class cognitive structures of the given stimulus words (concepts).

4.3.5 Graphic Map of the Pupils' Cognitive Structures:

Giving pupils the stimulus words and allocating each pupil raw scores is not the end of the story of the word association technique. The teacher may want to summarise the pupils' responses by drawing a map showing their cognitive structures. There are two methods commonly used to draw a map of pupils' cognitive structures; the relatedness coefficient and frequency responses. Both methods use the Waern (1972) method of drawing the cognitive structures of pupils' minds.

4.3.5.1 Relatedness Coefficient Method:

In this method, the relatedness coefficient should be first calculated for each pair of the stimulus words for each pupil using the formula above. Then, a mean relatedness coefficient should be calculated for the whole class for each stimulus words pair. To present the results of the calculations in a graphic manner, the strongest relatedness coefficient greater than certain value of RC between any two stimulus words is shown (drawn) first. Secondly, the coefficient is relaxed to a new value and new stimulus word(s) are attached to the existing ones. This will continue until all the stimulus words are joined in the picture and the complete map of the pupils' cognitive structures appears. The line thicknesses might be used to indicate the relative strengths of the associations.

This method has two limitations:

- It shows only the association between the stimulus words and ignores the responses.
- It takes a lot of effort and long time to calculate the relatedness coefficient for each pupil for each pair of the stimulus words.

4.3.5.2 Frequency Responses Method:

This method has been used by some researchers such as Bahar *et al.* (1999) and Bahar (1999). The first step in this method is to count, for the class as a whole, the type and number of responses for each stimulus word. The responses, which were counted, were taken to be 'valid' if they were meaningful and acceptable in terms of the assessed topic.

After the count was made, the word(s), which were mentioned more than certain number of times in response to a stimulus word(s) was shown (drawn) first. Then, the criterion will be lowered to a new frequency of mention and a new stimulus word(s) and responses will join the picture. This procedure will continue until all the stimulus words join the picture and the complete map of pupils' cognitive structures can be generated. It is not necessary to start with one stimulus word but it may be that there are several stimulus words which have high frequency responses.

The frequency responses method yields a complex but informative diagram. This method does not take into account the rank order of the responses.

4.3.6 Word Association Technique and Assessment in Science Education:

Johnstone *et al.* (1985) used WAT to map the cognitive structures of the pupils of some chemistry concepts. In addition, they used pupils' responses as distractors in the achievement testing (multiple choice questions). The results demonstrated that WAT does reflect the underlying fine structure of the cognitive structure. They found that if the existing concepts are firmly linked in the cognitive structure, then the facility values of the questions testing these concepts are high. The same results were reached by Cachapuz *et al.* (1987) in their study. They found that if the relationships did not appear in the associative structures from the WAT, then pupils tended not to be able to solve achievement test items which required the concept relations.

4.3.7 Conclusions of Word Association Technique:

The conclusions are that this technique can be used as a tool of teaching, learning and diagnostic assessment. The teacher may use it during the learning and teaching process with normal classroom tests as a rich source of data to guide him in his teaching. Also it can be used prior the formal exam as an objective source of distractors for MCQs.

CHAPTER FIVE: FIELD RESEARCH AND RESULTS (2) - PART (A)

CHAPTER FIVE: FIELD RESEARCH AND RESULTS (2) – PART (A)

5.1 Introduction:

The aims of this chapter are to describe the methods, which were used to investigate the effect of some psychological factors on pupils' responses in some fixed and open response methods of assessment and to present the findings of these effects.

5.2 Cognitive Styles and Assessment Methods:

This section examines the effect of two cognitive styles, field dependent – independent and convergent – divergent, on pupils' responses in some fixed and open response methods of assessment.

5.2.1 Field Dependent - Independent Cognitive Style and Assessment Methods:

5.2.1.1 The Aims of the Research:

The aims of this part of the research are to examine whether:

- There is a correlation between pupils' performances in structural communication grids and their achievements in field dependent – independent cognitive style test.
- There is a difference between field - dependent learners and field - independent learners in their performance on fixed and open response questions in school chemistry. The fixed response questions were based on the grid questions.

5.2.1.2 The Research Hypotheses:

Three hypotheses were tested in this part of the research. These are as follows:

Hypothesis 5.1: There is no correlation between pupils' performances in a field dependent - independent test and their achievement in structural communication grids (SCGs).

Hypothesis 5.2: Field - independent pupils are expected to perform better than field - dependent pupils in fixed response questions based on the grid questions partly because field - independent pupils tend to have high working memory capacity and partly because they will not distracted by the grid.

Hypothesis 5.3: There is no difference in performance between field - dependent pupils and field - independent pupils in open response questions.

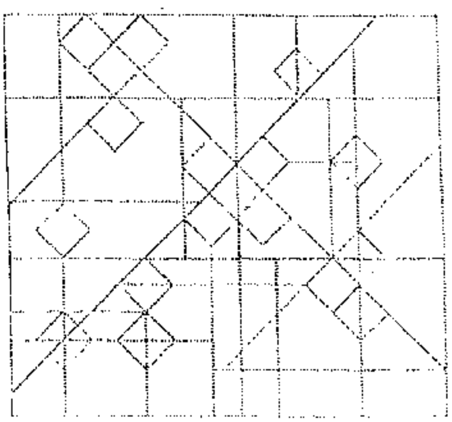
5.2.1.3 Description of Field Dependent - Independent Test:

i. Introduction:

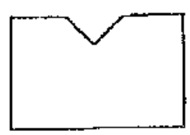
Field dependent - independent test (also called Hidden Figure Test (HFT)), which was used in this research, is based on the work of Witkin and his colleagues (Witkin *et al.*, 1981). It has undergone further design and standardised at the Centre for Science Education - University of Glasgow. In this test, the subject is presented with an item in a simple geometric design which is contained within a complex organized field (i.e. complex design). The test contains both the single item (simple geometric design) and noise (complex design). The surrounding organized field has influenced the person's perception of an item within it. Therefore, the general purpose of the test is to classify the sample of the research into field – dependent pupils who find difficulty or are unable to discriminate a required item from its context and field – independent pupils who are able to discriminate the required item from its context. In other words, those who can separate the '**signal**' from surrounding distracting '**noise**'.

HFT requires students to find a simple shape hidden in a complex pattern of lines. The current test consists of twenty complex patterns of lines apart from another two figures used as examples. When pupils find the shape, they have to outline it using a pen. One mark has been allocated for each correct tracing of the simple shape within a complex one. That means, the maximum score a pupil can get is 20.

Example of Field Dependent – Independent Test:



Complex Design



Simple Shape to be Traced

ii. Conditions of Applying the Hidden Figure Test:

There are some rules that should be considered when HFT is administered.

- The traced simple shape should be the same size, and have the same proportions and face the same direction as when it is appears alone in the specimen.
- In each complex pattern there are several simple shapes hidden but the required simple shape appears only once in the complex pattern. The pupil should trace the required shape only and nothing else.
- A pupil is not allowed to use a ruler or any other means to measure the size of the simple shape in the complex pattern.
- Pupils are given 15 minutes to complete tracing the real exercise plus 5 minutes to read the instructions and try the examples.

5.2.1.4 The Sample:

The sample consisted of ninety two (92) pupils (Age 15-16). They were selected from one secondary school in the Central Region of Scotland. All pupils were taking the Standard - Grade chemistry course.

There is no one certain criterion, in the literature, which has been used to classify the sample into the three categories of field dependent - independent cognitive style. For example, Ziane (1990) classified his sample as follows:

Field Dependence: students who scored from 2 to 6 in field dependent – independent test.

Field Intermediate: students who scored from 7 to 12 in field dependent – independent test.

Field Independence: students who scored 13 and upwards in field dependent – independent test.

On the other hand, Al – Naeme (1991) and Hasson (1988) used half standard deviation (0.50) to classify their sample, whereas Bahar (1999) used quarter standard deviation (0.25) to classify his sample. The researcher used 0.40 of the standard deviation to classify pupils' sample into the three categories of field dependent – independent cognitive style. This value seemed to be the best cut off point in which there was no bias towards any category.

According to this, 0.40 of the standard deviation on each side of the mean has been used to assign pupils into the Field – Intermediate category. Above that cut – off pupils were classified as - Field Independent and below as Field – Dependent (Figure 5.1).

The following steps show how pupils' were classified into the three categories of field dependence – independence.

<u>STEP ONE:</u>

- The Mean Score of the Sample = 8.46
- Standard Deviation (SD) = 3.95
- 0.40 of the Standard Deviation = 1.58

<u>STEP TWO:</u>

Field - dependence (FD) = 8.46 - 0.40SD = 8.46 - 1.58 = 6.88

→ FD = 6.88 and below

Field - independence (FI) = 8.46 + 0.40SD = 8.46 + 1.58 = 10.04

→ FI=10.04 and above

Field - Intermediate (FInter.) = In between (6.88 ↔ 10.04)

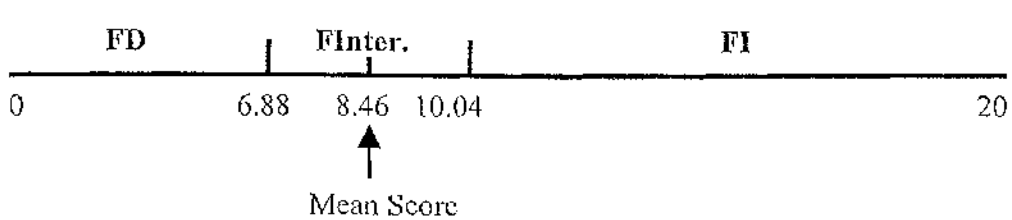


Figure 5.1: The classification of the sample into the three categories of field dependent – independent cognitive style

According to above classification, twenty nine pupils were classified as field - dependent, thirty five as field - independent and the remaining as field - intermediate (Table 5.1).

Table 5.1: Number of pupils in each category of field dependent - independent cognitive style

Category	Number of Pupils	Percentage (%)
Field dependence (FD)	29	32
Field Intermediate (F Inter.)	28	30
Field Independence (FI)	35	38
Total	**92**	**100**

5.2.1.5 Standard - Grade Chemistry Preliminary Exam:

This exam is usually held in December or January. It is divided into two parts. Part one consists of fixed response questions (all in the grid form) and the pupil answers these questions by drawing a circle around the appropriate letter(s) on the grid sheet provided. Part two consists of open response questions and the pupil answers these questions by writing his answers on the answer sheets provided. Pupils' scores in these types of questions were used for the research purposes.

5.2.2 Convergent - Divergent Cognitive Style and Assessment Methods:

5.2.2.1 The Aim of the Research:

The aim of this part of the research is to investigate the effect of convergent - divergent cognitive style on pupils' performances in two types of fixed response questions (MCQs and SCGs) in Higher - Grade chemistry.

5.2.2.2 The Research Hypotheses:

This research is intended to test the following hypotheses:

Hypothesis 5.4: There is no correlation between pupils' performances in convergent - divergent test and their achievement in fixed response questions.

Hypothesis 5.5: The performance of convergent pupils is expected to be better than divergent pupils in the MCQs because, in MCQs, they are rewarded for convergent thinking leading to a unique answer.

Hypothesis 5.6: Divergent pupils are expected to perform better than convergent pupils in SCGs as is supported from previous study (i.e. Bahar's (1999) study).

5.2.2.3 Description of the Convergent - Divergent Test:

i. Introduction:

This test has been designed and modified by Al - Naeme (1991) from Hudson's original test (Hudson, 1966). The current test consists of six sub - tests. The pupil is required to do different tasks in each sub - test. The time allocated to finish each test is limited. In each sub - test, spaces are provided for answering the required task. One mark has been assigned to each correct space. The maximum score a pupil can get is 98. Pupils are allowed to move to the next test only when the time of the previous one is up.

ii. Descriptions of the Six Sub - Tests:

Sub - test one:

This sub - test asks pupils to think of words having meanings which are the same as or similar to a given word. Strong, clear and dark are the three words, which have been given to pupils to find synonyms.

Sub - test two:

In this sub - test, pupils are given four different words. Then, they are asked to write as many sentences as they can using these four words and any other words chosen. The words must be used in the forms that are given. Two different sets of words are presented for the pupil to write sentences using them in the test.

Sub - test three:

In this sub - test, pupils are asked to think up and draw a number of different symbols that could be used to stand for certain words or ideas. Energy, happiness, technology and silence are the four words pupils are asked to draw symbols about.

Sub - test four:

This sub - test aims to test how many things that pupils can think of that are alike in some way. Pupils are asked to write as many things as possible that are 'round' or that are round more often than any other shape.

Sub - test five:

In this sub - test, pupils are asked to think and write as rapidly as they can about the words that begin with one letter and end with another. Pupils are asked to write the words that begin with letter 'G' and end with letter 'T'. Names of the people or places are not allowed.

Sub - test six:

This sub - test aims to test the ability of pupils to think and write as many ideas as they can think of about a topic. It might be that some of ideas related to the topic are not important to pupils but does not matter. 'Working in laboratories' is the topic that pupils are asked to write as many ideas as they can.

5.2.2.4 The Sample:

The sample was ninety four (94) pupils (Age 16-17) from four high schools in the Central Region of Scotland (Table 5.2). These pupils were taking Higher - Grade course in chemistry.

Table 5.2: Number of pupils taking convergent - divergent test from the four high schools

School	Number of Pupils	Percentage (%)
School One	33	35
School Two	21	22
School Three	13	14
School Four	27	29
Total	**94**	**100**

Again and the same as in field dependent – independent cognitive style, there is no one certain cut off point taken to assign the pupils into the three main categories of convergent – divergent cognitive style. Al – Naeme (1991), for example, used the mean score as a crucial point to classify his sample into the three categories of convergent – divergent cognitive style. He ignored the all – rounders and classified his sample into only convergers and divergers. However, Bahar's (1999) classification did not ignore the all – rounders. He used quarter standard deviation (0.25) to assign his sample into convergers, all – rounders and divergers. The current research used 0.40 of the standard deviation as the crucial point to classify the pupils into the three main categories of convergent - divergent cognitive style (Figure 5.2). The following steps show how pupils were classified into the three categories of convergence – divergence:

STEP ONE:

- The Mean Score of the Sample = 46.2
- Standard Deviation (SD) = 10.1
- 0.40 of the Standard Deviation = 4.04

STEP TWO:

Convergers = 46.2 – 0.40 SD = 46.2 – 4.04 = 42.2

→ Convergers = 42.2 and below

Divergers = 46.2 + 0.40SD = 46.2 + 4.04 = 50.2

→ Divergers = 50.2 and above

All - Rounders = In between (42.2 ↔ 50.2)

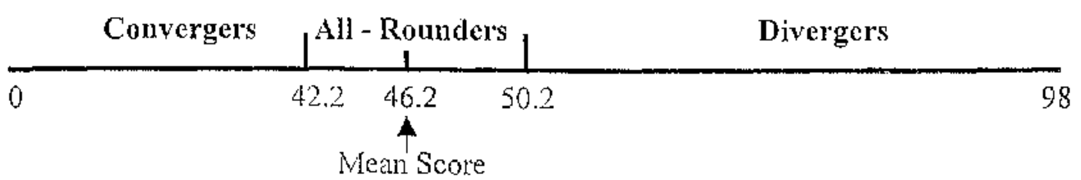


Figure 5.2: The classification of the sample into the three categories of convergent – divergent cognitive style

From the classification above, thirty four pupils were classified as convergers, twenty nine pupils as all - rounders and the remaining thirty one pupils were classified as divergers (Table 5.3).

Table 5.3: Number of pupils in each category of the convergent - divergent cognitive style

Category	Number of Pupils	Percentage (%)
Convergers	34	36
All - Rounders	29	31
Divergers	31	33
Total	94	100

5.2.2.5 Chemistry Preliminary Exam for the Higher - Grade Pupils:

Higher - Grade pupils take a preliminary exam usually in chemistry and other science disciplines in December or January. The exam consists of two types of questions: fixed response and open response. The fixed response questions are divided into multiple choice questions (MCQs) and structural communication grids (SCGs). Forty marks were allocated for the MCQs and twenty marks for the SCGs. The pupils' scores in fixed response questions were used for testing the hypotheses.

5.3 The Results of Applying the Two Cognitive Styles on the Assessment Methods:

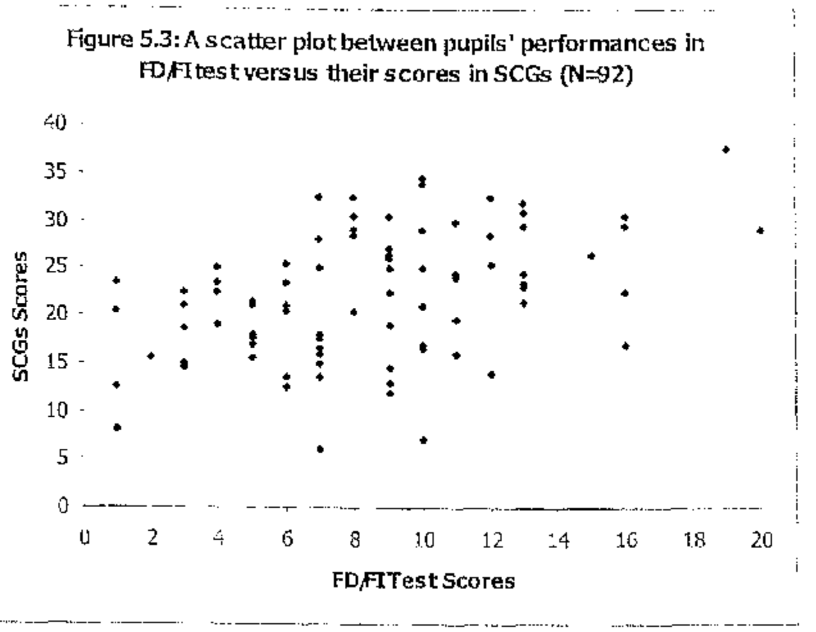
After we have seen the methodology, which used to investigate the effect of field dependent – independent and convergent – divergent cognitive styles in section 5.2, the time now to see the results of this investigations.

5.3.1 Field Dependent - Independent Cognitive Style and Assessment Methods:

In the current chapter, section 5.2.1.2, three research hypotheses were proposed. Let us see which of these hypotheses can be accepted and which can be rejected.

5.3.1.1 Field Dependent - Independent Cognitive Style and SCGs:

As we know previously (Chapter Four, section 4.1.2.1) that field dependent – independent cognitive style appears in the majority of psychological domains, providing a coherent and consistent description of individual behaviour differences. Consequently, we expect that there is a correlation between pupils' performances in field dependent-independent test and their achievement in structural communication grids (SCGs). To find out this correlation, Pearson Product Moment Correlation Coefficient was applied and the results revealed that there is a positive correlation ($r = 0.419$, Degrees of freedom 90/ two - tailed test) between the two tests (Figure 5.3).



Therefore, the null hypothesis (hypothesis 5.1); that is '*there is no correlation between pupils' performances in FD/FI test and their achievement in SCGs*' can be rejected at the 0.001 level. The conclusion of this is that, field - independent pupils had higher scores than field - dependent pupils in the SCGs.

5.3.1.2 Field Dependence - Independence and Fixed and Open Response Questions:

In this investigation the results showed that the field - independent pupils perform better than field - dependent pupils in fixed (in the grid form) and open response questions (Table 5.4).

Table 5.4: The mean score and standard deviation of fixed response and open response questions versus field dependence – independence categories

	Types of Questions			
	Fixed Response (SCGs) *		**Open Response •**	
FD/DI	**Mean Score**	**SD**	**Mean Score**	**SD**
FD	19.1	4.30	36.9	9.53
FInter.	22.2	7.44	45.1	15.24
FI	24.4	6.94	47.6	16.1

Note:-

* *Marks out of 40.*

• *Marks out of 80.*

As is shown in Table 5.4, the mean score of field - independent pupils is larger than the mean score of field - dependent pupils in both the fixed response and open response questions. To test if this difference is statistically significant or not, t - test was used. The resultant value of the t - test showed that the difference is statistically significant at 1% level for fixed response questions (Degrees of freedom = 62 /two - tailed test). Therefore, hypothesis **5.2** which states that '*Field - independent pupils are expected to perform better than field - dependent pupils in fixed response questions based on the grid questions partly because field - independent pupils tend to have high working memory capacity and partly because they will not distracted by the grids*' can be accepted at 1% level.

The same results were found for the open response questions. From the same table, it can be seen that the mean score of field - independent pupils is higher than the mean score of field - dependent pupils in the open response questions. The difference in the mean score between the two groups was found statistically significant at 1% level when the t - test was applied (Degrees of freedom = 62/ two - tailed test). Therefore, hypothesis **5.3**, that is '*there is no difference in performance between field – dependent pupils and field – independent pupils in open response questions*' can be rejected at 1% level.

5.3.2 Convergent - Divergent Cognitive Style and Assessment Methods:

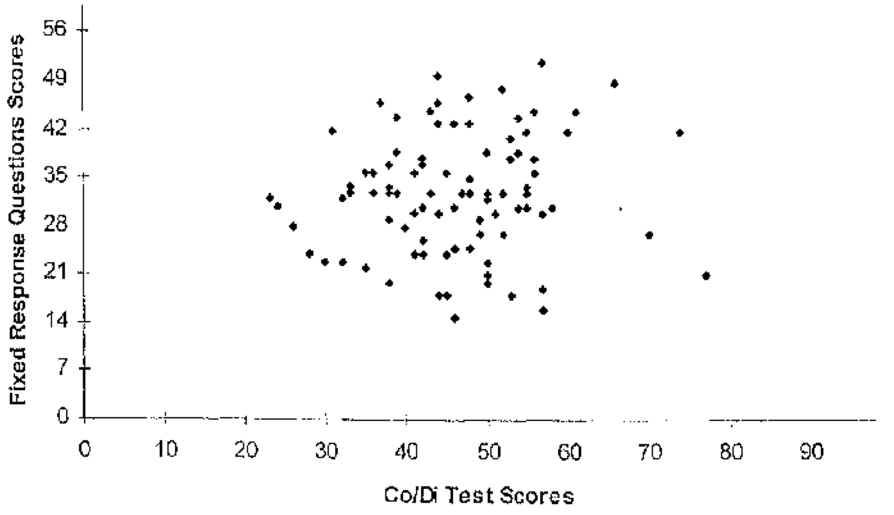
We saw in the previous section (section 5.3.1) how field dependent - independent cognitive style has an effect on pupils' performance in some types of assessment methods. However, will this be the same for the convergent - divergent cognitive style?

5.3.2.1 Pupils' Performances in Co/Di test and their Achievement in Fixed Response Questions:

Two fixed response methods of assessment (MCQs and SCGs) were used to arrive at a conclusion if pupils' achievement in fixed response is related to their performance in the convergent - divergent test. Pupils' scores in higher – grade chemistry preliminary exam in MCQs and SCGs were correlated with their scores in convergence - divergence test. The results (Figure 5.4) showed that there is a small positive correlation between the two tests ($r = 0.116$, Degrees of freedom = 92/ two - tailed test). However, this correlation was found to be statistically not significant at the acceptable levels (5% or 1%). Therefore, the null hypothesis (hypothesis **5.4**) which proposes that '*there is no* *correlation between pupils' performances in convergent – divergent test and their achievement in fixed response questions'* can be accepted.

The conclusion, which has emerged from the above results, revealed that pupils' achievements in fixed response questions are not related to their degree of divergency.

Figure 5.4: A scatter plot between pupils' scores in Co/Di test versus their scores in fixed response questions (N = 94)



Two other hypotheses were suggested to be tested in this part of the research. These are hypothesis 5.5 which asserts that '*the performance of convergent pupils is expected to be better than divergent pupils in the MCQs because, in MCQs, they are rewarded for convergent thinking leading to a unique answer'* and hypothesis 5.6 which states that '*divergent pupils are expected to perform better than convergent pupils in SCGs as is supported from previous study (Bahar's (1999) study)'*.

To reveal and examine the above two hypotheses, the mean scores of both groups (convergers and divergers) in MCQs and SCGs were calculated and compared (Table 5.5).

Table 5.5: The mean score and standard deviation of each category of convergent - divergent cognitive style in MCQs and SCGs

	Types of Questions			
	MCQs *		SCGs •	
Category	**Mean**	**SD**	**Mean**	**SD**
Convergers	21.5	5.14	10.5	3.23
All - Rounders	22.1	7.41	10.5	3.64
Divergers	24.1	6.82	10.9	4.26

Note:-

* *Marks out of 40.*

• *Marks out of 20.*

It can be seen from Table 5.5 that the mean score of divergent pupils is greater than the mean score of convergent pupils in MCQs. The t – test was applied to find out if this difference is statistically significant or not. The value of the t – test showed that this difference is not statistically significant at 1% or 5% levels. Therefore, hypothesis **5.5** can be rejected.

Also we can see from the same table that there is no real difference between the mean score of convergent pupils and the mean score of divergent pupils in SCGs. As a result of this, hypothesis 5.6 can be rejected.

CHAPTER FIVE: FIELD RESEARCH AND RESULTS (2) - PART (B)

CHAPTER FIVE: FIELD RESEARCH AND RESULTS (2) – PART (B)

This part is written to continue from the previous part on the effect of some psychological factors on pupils' responses in some fixed and open response methods of assessment.

5.4 The Relationship between Pupils' Knowledge Structures and Multiple Choice Test Structure:

Word association test (WAT) has been used for a long time as a tool for diagnostic purposes to investigate pupils' learning. However, in the current research, the use of this method will be shifted from a learning purpose to an assessment purpose.

5.4.1 The Aims of the Research:

This part of the research aims to:

- Map the pupils' knowledge structures of the given eight chemical concepts using a word association test.
- Investigate the effect of the positions of the options, in the pupils' knowledge structures, in their selections in the multiple choice test.

5.4.2 The Research Hypothesis:

The following hypothesis was tested in this part of the research:

Hypothesis 5.7: Pupils are expected to select the option which is closer, in the pupils' knowledge structures, to the concept on which the stem of the question is based.

5.4.3 Descriptions of Chemistry Word Association and Multiple Choice Tests:

5.4.3.1 Word Association Test (WAT) and the Graphic Map:

The given key words in chemistry WAT were provided by the principal teacher of chemistry. These seemed the most important ideas which had been taught recently. These were Chemical Reaction (as an example), Periodic Table, Element, Compound, Atom, Metal, Non - Metal, Molecules and Matter (Examples are shown in Appendix 6.1).

In the *first page* of the WAT the following instructions are given to pupils:

> "When you hear or see a word, it often makes you think of other words. In this study we would like to find out what other words are brought to your mind by some words in chemistry.
>
> On each page you will find a key word written many times. Say the word to yourself, and then as quickly as possible write the first word comes to your mind in space Number 1. And then continue to write in other spaces other associated words come to your mind.
>
> Continue in this way until you are told to turn the next page. There are no right answers. Write as quickly as possible since you are allowed only **one minute (1 MINUTE)** for each page".

The *second page* of the test is an example of responses to the stimulus word 'Chemical Reaction'. The time allowed for each stimulus word was varied among the researchers. Some researchers such as Kempa and Nicholls (1983) and Bahar (1999) allowed thirty seconds for each stimulus word. It is due to the fact that thirty seconds had been established as an optimum time span during pre - tests (Cachapuz and Maskill, 1987). On the other hand, some researchers such as Johnson, (1967; 1969) and Shavelson (1972) allowed one minute for each stimulus word with pupils in high school. The researcher allowed one minute for each stimulus word because:

- Most of the researches on WAT done at high school allowed one minute.
- Pupils, who took the test, were so young (12 years old) that they were slow in their writing. Therefore, it was important to give them enough time.

After the WAT was marked and the number and types of responses for each stimulus word was counted, the graphic map for the eight chemical concepts was produced (Some of the Response and their Frequencies are given in Appendix 6.2). Then, this map was used to write and construct the multiple choice test.

5.4.3.2 Multiple Choice Test:

The multiple choice test was constructed from the responses which appeared in the graphic map of the pupils' knowledge structures (Figure 5.5). The test was designed in such away that the options for each item were chosen from different positions in the pupils' knowledge structures. The multiple choice test consisted of twelve (12) items (Appendix 6.3). The test was checked twice; firstly by the specialists in chemistry from the Centre for Science Education - University of Glasgow and secondly by the principal and class teachers of chemistry of the school. The multiple choice test was administered to pupils one month after applying the word association test. One mark was assigned for each correct answer. The maximum score a pupil can get is 12 (i.e. twelve multiple choice items).

Examples of the Chemistry Multiple Choice Test:

The following are examples of the chemistry multiple choice questions.

Example One:

In this example, the stem of the question is based on the key word 'element' and the options were selected from different positions in the graphic map of the pupils' knowledge structures (see Figure 5.5, page 169).

Which of the following elements has the chemical symbol of 'Ag'?

a. Tin.
b. Copper. → Further away from the key word 'element'
c. Silver.
d. Oxygen. → Close to the key word 'element'

In this example, we expect that the facility value of option **d** (oxygen) will be higher than the remaining options. In other words, we predict that most of the pupils will choose this option and less pupils will choose options **a**, **b** and **c**.

Example Two:

In this example, the stem of the question is based on the key word 'element'. Again, the options were chosen from different positions in the graphic map of pupils' knowledge structures (see Figure 5.5, page 169).

Which one of the following elements helps things to burn?

a. Oxygen.
b. Hydrogen. → Close to the key word 'element'
c. Copper.
d. Aluminum. → Further away from the key word 'element'

We predict that most pupils will choice options **a** and **b** because they are located closer to the key word 'element' compare to options **c** and **d**. The facility value of the former options will be higher than the facility value of the latter options.

Example Three:

In this example, the stem of the question is based on the word 'small'. This word is not one of the key words, however. It is one of the responses given by pupils (see Figure 5.5, page 169).

Which one of the following is the smallest among the others?

a. Molecules. ⟶ Close to the word 'small'
b. Compounds. ⟶ Further away from the word 'small'
c. Atoms. ⟶ Close to the word 'small'
d. Elements. ⟶ Further away from the word 'small'

Our expectation in this example is that most pupils will select options **a** and **c**. This is due to the fact that these options are located closer to the word 'small' in the graphic map of the pupils' knowledge structures. On the other hand, options **b** and **d** are less likely to be selected because they are located further way from the word 'small' in the pupils' knowledge structures.

Example Four:

The fourth example is based on the key word 'molecules'. The four options of the item were selected from different positions in the graphic map of pupils' knowledge structures (see Figure 5.5, page 169).

In which of the following are the molecules far apart from each other?

a. Wood. ⟶ Further away from the key word 'molecules'
b. Air. ⟶ Close to the key word 'molecules'
c. Water. ⟶ Close to the key word 'molecules'
d. Glass. ⟶ Further away from the key word 'molecules'

In this example, options **b** and **c** are expected to have high facility value (selected by most of the pupils) because they are located closer to the key word 'molecules'. The remaining options (**a** and **d**) are expected to have low facility value (selected by few pupils) because they are located further away from the key word 'molecules' in the graphic map of pupils' knowledge structures.

5.4.4 The Sample:

The sample was first year secondary school pupils (Age 12-13) from one high school in the Central Region of Scotland who were taking the chemistry part of general science. One hundred (100) pupils did the WAT, whereas sixty one (61) pupils did the multiple choice test. Both the word association and multiple choice tests were administered under exam conditions.

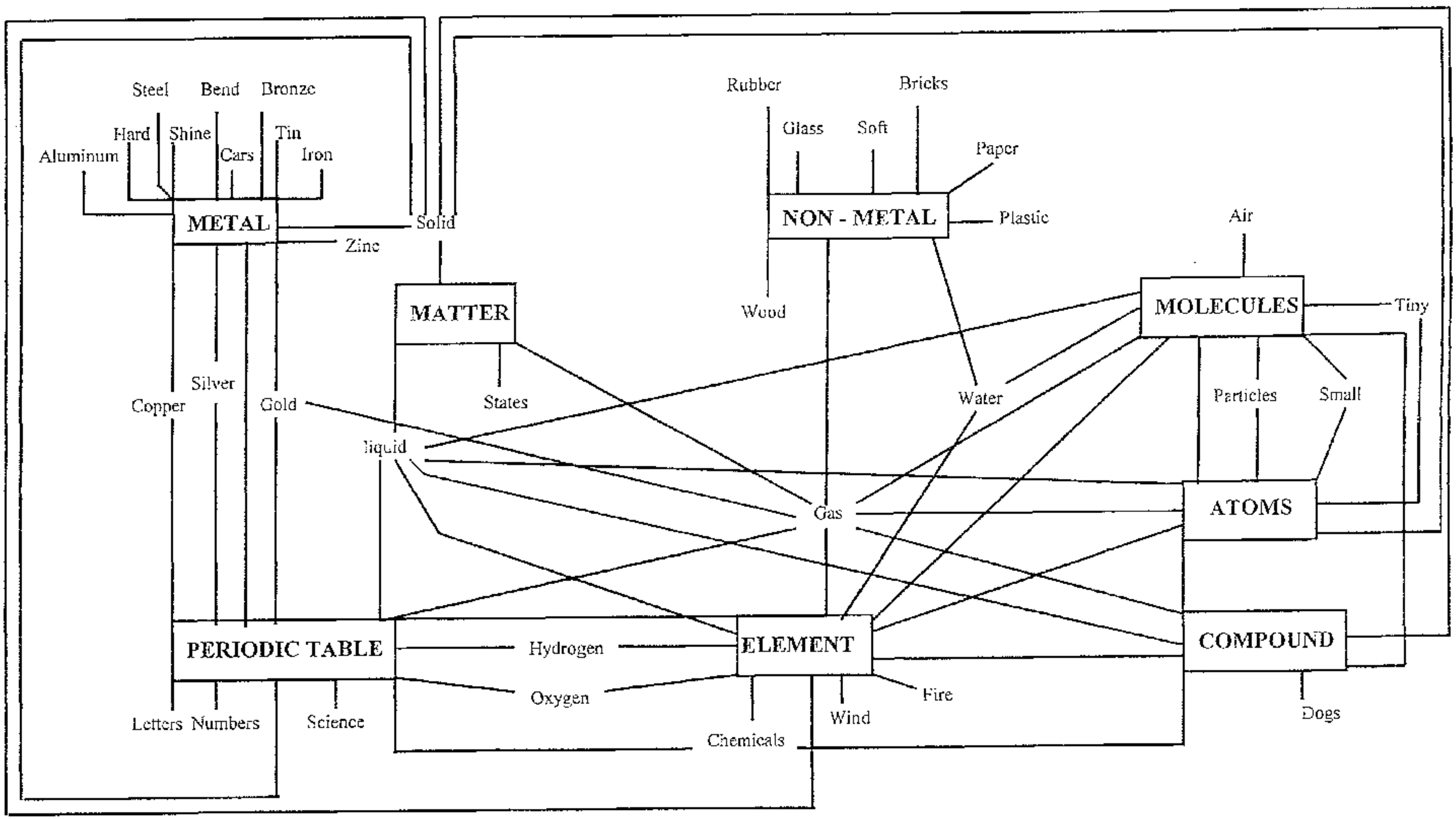


Figure 5.5: The graphic map of the eight chemical key words of the school level

5.5 The Results of the Relationship between Pupils' Knowledge Structures and Multiple Choice Test Structure:

5.5.1 Introduction:

It is not easy to construct a multiple choice test based on the responses which appeared in the graphic map of pupils' knowledge structures. These are due to the following constraints:

- Not all the responses, which were given by pupils in WAT. are in the same form of the language (i.e. mixture of verbs, nouns, adjectives, adverbs.........etc.).
- Some responses were not shown in the graphic map because of their low frequency unless a very low cut - off point is taken.

In spite of the above obstacles, we still can reach a conclusion about the relationship between pupils' knowledge structures and the multiple choice test structure.

5.5.2 The Facility Value of Each Option of Each Item of the Multiple Choice Test:

The facility value of each option of each item of the multiple choice test was calculated and presented in Table 5.6. By calculating the facility value of each option, we will be able to understand the effect of the positions of the options, in the pupils' knowledge structures, in their selections.

Table 5.6: The facility value (F.V) of the key answer and different distractors of the chemistry multiple choice test

	Facility Value				
Item Number	Key Answer	Distractor (1)	Distractor (2)	Distractor (3)	Omitting
1	c) 0.93	a) 0.05	b) 0.02	d) -	-
2	b) 0.74	a) 0.07	c) 0.13	d) 0.05	0.02
3	c) 0.66	a) 0.05	b) 0.10	d) 0.16	0.03
4	a) 0.74	b) 0.20	c) -	d) 0.07	-
5	c) 0.67	a) 0.02	b) -	d) 0.31	-
6	d) 0.52	a) 0.10	b) 0.33	c) 0.03	0.02
7	c) 0.64	a) 0.10	b) 0.16	d) 0.10	-
8	a) 0.62	b) 0.25	c) 0.02	d) 0.08	0.03
9	b) 0.34	a) 0.28	c) 0.08	d) 0.30	-
10	d) 0.18	a) 0.38	b) 0.20	c) 0.25	-
11	c) 0.48	a) 0.34	b) 0.08	d) 0.07	0.03
12	b) 0.69	a) 0.08	c) 0.13	d) 0.05	0.05

Generally speaking, in the relationship between pupils' knowledge structures and the multiple choice test structure is that the option in the question which is closer to the concept on which the stem of the question is based is more likely to be selected than other options.

For example (from Table 5.6), the stem of item one is based on the key word '**element**'. Option **c** (hydrogen) was selected more than options **a** and **b** (zinc and gold respectively). The facility value of option **c** is 0.93, whereas the facility value of option **a** is 0.05 and of option **b** is 0.02. Option **c** (the key answer) is located closer to the key word '**element**' in the pupils' knowledge structures compared to the remaining options.

The second example is item four. The stem of this item (example two of section 5.4.3.2) is also based on the key word '**element**'. In this item, the two options, which we predicted to be selected by most of the pupils, were found matches with our prediction. The facility value of option **a**; the key answer (oxygen) is 0.74 and of option **b**; hydrogen is 0.20. These two options, as stated above, are located closer to the key word '**element**' in the pupils' knowledge structures. The remaining options (c and d) had low effect in distracting pupils and they are located further away from the key word in the pupils' knowledge structures.

The third example, which may support the hypothesis, is item eight. In this item, option **a** (the key answer) has high facility value compared to the rest of the options. This option (hydrogen) is located very close to the key word '**element**' in the pupils' knowledge structures. Another option, which attracts some pupils, is option **b** (oxygen). The facility value of this option is 0.25. It is also like option **a** located closer to the key word '**element**' in the pupils' knowledge structures. The remaining options **c** (gold), **d** (copper) are located further away from the key word in the pupils' knowledge structures. These options were found to have a very low effect in distracting pupils.

Option **c** (atoms) is the key answer of item eleven (example three of section 5.4.3.2). In this item, we predicted that most of the pupils will select options **a** and **c**. The result matches our expectation. Option **a** (molecules) and option **c** (atoms) were selected more among the remaining options (F.V of option **a** = 0.48 and = 0.34 of the option **c**). These options are located very close to the concept '**small**', in the pupils' knowledge structures, where the stem of the item is based. Options **b** (compounds) and **d** (elements) were found to have very small facility value (F.V of option **b** = 0.08 and of option **d** = 0.07).

The final example, which is worth looking at, is item twelve (example four of section 5.4.3.2). The stem of this item is based on the key word '**molecules**'. Option **b** (air), which is the key answer, has the largest facility value among the options (F.V = 0.69). This option was predicted to have high facility value as it is located very close to the key word in the pupils' knowledge structures. Option **c** (water) is another option that attracted some pupils. This option was also predicted to have high facility value because it is located close to the key word in the pupils' knowledge structures (F.V = 0.13). The remaining two options (**a** (wood) and **d** (glass)) had no effect in distracting pupils. They are located further away from the key answer in the pupils' knowledge structures.

However, there are some exceptions from the above observation. For example, we expected that the facility value of option **d** (oxygen) of item three (example one of section 5.4.3.2), will be higher than the facility value of the remaining options. However, the result revealed that it is not the case. It was found that the facility value of this option was smaller than the facility value of option **c** (the key answer). Option **c** is located further away from the key word '**element**' in the pupils' knowledge structures.

Item five is another example of this expectation. In this example, the stem of the question is based on the key word ' **non – metal**'. We predicted that option **d** (paper) will be selected by most of the pupils (higher facility value) because it is located closer to the key word in the pupils' knowledge structures. Nonetheless, option **c** (oxygen = the key answer) appeared to strongly attract many pupils. The facility value of this option is higher than the facility value of option **d** (F.V of option **c** = 0.67 and of option **d** = 0.31). Option **c** is located further away from the key word in the pupils' knowledge structures.

One possible explanation of the above expectations is that it may be because the multiple choice test was administered to pupils one month after applying the word association test. That means, pupils may forget some parts of the knowledge, which were tested by word association test. The other explanation may be that some parts, in this area, were not given too much emphasis by the class teachers. The final explanation is related to the key answer. The option, which is the key answer, looks like having an important part in pupils' selections.

Consequently, hypothesis 5.7 (i.e. *pupils are expected to select the option which is closer, in the pupils' knowledge structures, to the concept on which the stem of the question is based)* can be accepted for most of the test items (Items: **1, 2, 4, 8, 10, 11** and **12**).

Notwithstanding, one important point which we would like to emphasis here, it is not necessary that the positions of the concepts (i.e. options) in a graphic map to the concept on which the stem of the question is based, is the only factor affects pupils' selections in the multiple choice test. It could be that there are other underlying factors which have also an importance in pupils' selections. The results of the current research showed the significance of the positions of the concepts (i.e. options), in a graphic map, in pupils' selections but it did not show the effects of other factors. As a result of this, the effects of other factors (if any) are needed to examine and this is another area where the new researcher can investigate.

5.5.3 The Implication of Knowing Pupils' Knowledge Structures in Constructing Fixed Response Questions:

Although applying word association test and drawing a graphic map of the pupils' knowledge structures is hard work and time consuming, it is a valuable technique in constructing fixed response questions such as MCQs. The assessor can use the graphic map as an objective source of choosing options from different positions of the pupils' knowledge structures. In Chapter Two, section 2.2.1 we stated that the words, which are chosen as options in objective testing, are subjective because they were chosen by the teacher. However, by applying WAT and drawing the graphic map the selections of the words become less subjective.

5.6 Pupils' Performances in WAT and their Achievement in MCQs:

The purposes of the word association test (WAT) in this part of the research were different from its purposes in the previous section in the way that we are not using it to draw a graphic map to find out the positions of the options in the pupils' knowledge structures (section 5.4). The purposes of the word association test in this part are listed below (section 5.6.1).

5.6.1 The Aims of the Research:

The aims of this part of the research are to:

- Map pupils' cognitive structures of eight biological concepts using WAT.
- Use WAT as an objective source for constructing MCQs in biology.
- Compare pupils' performance on WAT and their achievement on the multiple choice test.
- Investigate the effect of introducing, into the multiple choice test, distractors which appeared in pupils' responses in WAT.

5.6.2 The Research Hypotheses:

The following hypotheses were tested in this part:

Hypothesis 5.8: There is no correlation between pupils' performances in WAT and their achievement in MCQs in secondary biology.

Hypothesis 5.9: Pupils who are more able to associate concepts in biology WAT are expected to perform better than those pupils who are less able to associate concepts in biology WAT in the multiple choice questions.

5.6.3 Descriptions of Biology Word Association and Multiple Choice Tests:

5.6.3.1 Word Association Test:

The given key words in the biology word association test were selected and provided by the principal teacher of biology of one of the schools. The eight key words are Living Thing, Animal, Plant, Photosynthesis. Energy, Carbon Dioxide, Water and Oxygen (Examples are shown in Appendix 7.1).

The same instructions given in the chemistry word association test (section 5.4.3.1) were also given in this test. But the words, which are brought to pupils' minds, should be words in biology.

The test booklet contains ten pages; one for the instructions, one for the example (the stimulus word *Meat*) and the remaining pages for the eight stimulus words. The stimulus word was presented to pupils at the top of each page and again ten times down the left hand side, next to the places where the responses were to be written. The reason behind this is to avoid pupils giving their responses to the previous response instead to the stimulus word. This is what educators called chained responses (Cachapuz *et al.*, 1987). This problem can be solved, as it was mentioned above, by writing the stimulus word every time down the left hand side of the paper so that pupil will be forced to give his responses to the stimulus word.

One minute was used for each stimulus word in the test. The researcher marked the test and one mark was allocated to each 'acceptable' response in a biology sense. The maximum score was 80 (eight stimulus words and ten spaces are provided for each stimulus word).

5.6.3.2 Multiple Choice Test:

The responses, which were given by pupils, to each stimulus word were used to design the multiple choice test (Some of the responses and their frequencies are given in Appendix 7.2).

The multiple choice test consisted of twelve items (Appendix 7.3). It was first checked by the specialists in biology from the Centre for Science Education - University of Glasgow and secondly by the principal and class teachers of biology of the school. In the MCQs, one mark was assigned for each correct answer. The maximum score was 12 (i.e. twelve multiple choice items).

Examples of the Biology Multiple Choice Test:

The following are examples of the biology multiple choice questions.

Example One:

In this example, we expect that option **a** (starch) has a higher facility value in the multiple choice test than the rest of the options. The reason is it has a higher response frequency in the word association test and at the same time the correct answer.

Iodine is used to test for ....

Option	Response frequency
a. Starch.	21
b. Protein.	14
c. Minerals.	1
d. Vitamins.	3

← The response frequencies of each option in WAT.

Example Two:

In this example, we predict that most of the pupils will select option **d** as it has a higher response frequency in WAT. Also we expect that few pupils will select options **a**, **b** and **c** because they have small response frequencies in WAT.

The main source of energy for the Earth is......

Option	Response frequency
a. The Sun.	47
b. The Moon.	1
c. Electricity.	6
d. Plants.	211

← The response frequencies of each option in WAT.

Example Three:

In this example, we predict that option **d** (oxygen) will be selected more among the remaining options. This is due to the fact that it has a high response frequency in WAT.

Which gas is released from leaves during the process of photosynthesis?

Option	Response frequency
a. Carbon Dioxide.	27
b. Hydrogen.	4
c. Nitrogen.	5
d. Oxygen.	78

← The response frequencies of each option in WAT.

5.6.4 The Sample:

The sample was drawn from one high school in the Central Region of Scotland. Ninety (90) pupils from the second secondary (Age 13-14) were tested. The multiple choice test was applied fifteen days after the WAT. Both the word association and multiple choice tests were administered to pupils under exam conditions.

5.7 The Results of Pupils' Performances in WAT and their Achievement in MCQs:

This section is going to present the findings of the effect of pupils' ability to associate words, in their minds, on their achievements in the multiple choice test.

5.7.1 Graphic Mapping of Pupils' Cognitive Structures for the Eight Biological Key Words:

One aim of this part of the research is to draw a map for pupils' cognitive structures round eight biological key words (Figure 5.6). The map was drawn after calculating the response frequencies of each stimulus words in the WAT (Appendix 7.2). The frequency response method was used to draw this map (Bahar, 1999; Bahar *et al.*, 1999). This method, as described in Chapter Four (section 4.3.5.2), yields a complex but informative diagram.

When we look to the graph as whole, it is clear how the eight key words are linked to each other and to their responses at the lowest cut off point (cut off = 20). Most of the linkages radiate from ***living things, animal, plant, water and oxygen*** key words and fewer linkages radiate from ***photosynthesis, energy and carbon dioxide*** key words. This could be explained in terms that pupils might know more about the former key words than the latter key words. This comes from the fact that pupils might be taught about the five key words since they were at the primary level.

However, looking at the graph in more detail we can see that strong association (cut off = 60) appeared between the living things and plant and humans. Another strong association appeared at the same cut off point between plant and photosynthesis and between oxygen and breathing. We predicted this kind of association as oxygen is always associated with breathing, photosynthesis is always associated with plants and living things associated with humans and plants.

Lowering the cut off point to 45, new associations emerged. The key word 'animal' appears at this cut off point. It associates with two types of animals; dogs and cats. The reason why key word 'animal' was found strongly associated with dogs and cats might be because most houses have either a dog or cat (common in daily life). Moreover, two more key words appear at this cut off point. These are 'water' which associates with drinking and 'energy' which associates with food. One word 'air' appears at this cut off point linked to the key word 'oxygen'.

The key word 'carbon dioxide' only appears in the map at cut off point 30. This indicates that there are no strong associations of this key word in the pupils' cognitive structures. Three responses were found associate with this key word. These are gas, breathing and air. Furthermore, new responses appear which link to the exiting key words. For examples, green and roses associate to the key word 'plant' and pigs, cows and hamster associate to the key word 'animal'.

The weak associations appear at the cut off point 20. At this point a weak association appears between the key words 'plant' and 'water'. Also, weak associations were found between the key word 'energy' and running and between the key word 'photosynthesis' and sun (for more examples, see the overlays (Figure 5.6)).

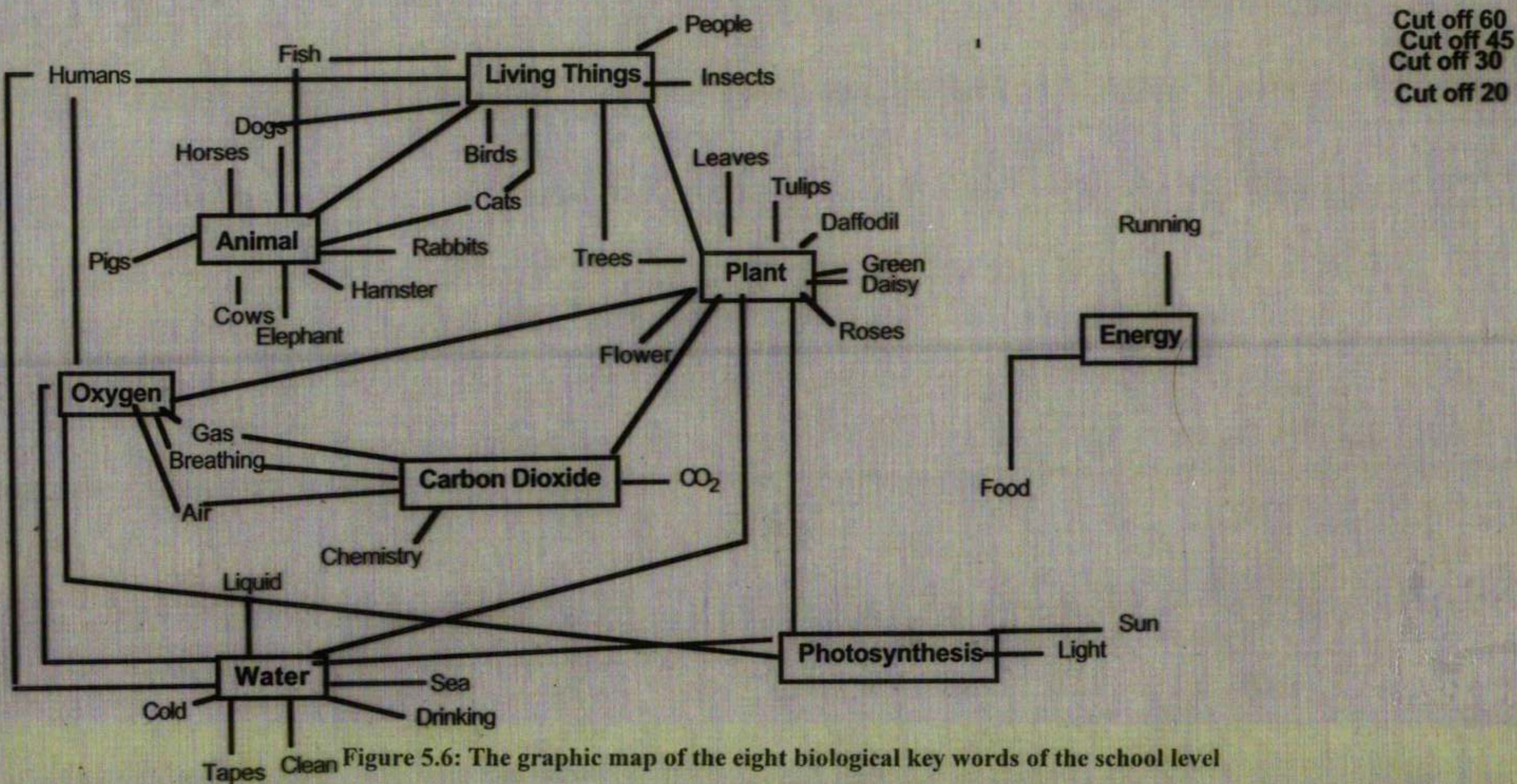


Figure 5.6: The graphic map of the eight biological key words of the school level

5.7.2 Pupils' Achievement in MCQs:

Pupils' mean score and standard deviation in WAT and MCQs were calculated and shown in Table 5.7 below.

Table 5.7: The mean score and standard deviation of biology word association test and multiple choice questions

	No. of Pupils	Mean (M)	SD
WAT *	90	50.7	13.1
MCQs •	90	8.44	2.34

Note:-

* *Marks out of 80.*

• *Marks out of 14.*

Hypothesis **5.8**, which states that '*there is no correlation between pupils' performances in WAT and their achievement in MCQs in secondary biology*', was tested in this part of the research. To reach a conclusion about this hypothesis, pupils' scores in WAT were correlated with their scores in biology MCQs using both the Spearman Rank Order Coefficient and Pearson Product Moment Coefficient. The results of these correlations are tabulated in Table 5.8 below.

Table 5.8: The correlation value between pupils' performances in biology word association test and their achievement in multiple choice test

	Correlation Value	Significant Level
Pearson	0.344	0.01
Spearman	0.387	0.05

From the above table (Table 5.8), it can be seen that there is a positive correlation between pupils' performances in WAT and their achievement in the MCQs. The above null hypothesis can be rejected at 5% level and the alternative hypothesis can be accepted.

5.7.3 Classification of the Sample in the WAT and their Achievement in MCQs:

According to pupils' mean score in the WAT (i.e. 50.7), pupils were classified into two groups; those who are more able to associate concepts in WAT (scores > 50.7) and those who are less able to associate concepts in WAT (scores < 50.7). The mean scores in WAT and MCQs for each group were calculated and presented in Table 5.9 below.

Table 5.9: The mean score and standard deviation of those pupils who are more able to associate words and those who are less able to associate words in WAT

		WAT		MCQs	
	No. of Pupils	(M)	SD	M	SD
More Able To Associate Words	49	60.7	6.51	9.06	2.29
Less Able To Associate Words	41	38.8	8.13	7.71	2.21

The results, which are shown in Table 5.9, matched with hypothesis **5.9**. This hypothesis predicts that pupils who are more able to associate concepts in WAT will perform better in MCQs than those pupils who are less able to associate concepts in the same test. The mean score of the former group is higher than mean score of the latter group in the multiple choice test. When t - test was applied, the difference of the mean scores of the two groups in MCQs was found statistically significant at 1% level (Degrees of freedom = 88/ two - tailed test). As a result, hypothesis **5.9** which states that '*pupils who are more able to associate concepts in WAT are expected to perform better than those pupils who are less able to associate concepts in biology WAT in MCQs*' can be accepted.

To find out how each group performed in each item of the multiple choice test, the facility value of each item for both groups was calculated and shown in Table 5.10 below.

Table 5.10: The facility value of each item of MCQs for those pupils who are more able to associate words and those who are less able to associate words in WAT

	Facility Value	
Item Number	**More Able to Associate Words**	**Less Able To Associate Words**
1	0.57	0.34
2	0.92	0.80
3	0.53	0.46
4	0.98	1.0
5	0.86	0.83
6	0.73	0.61
7	0.94	0.93
8	0.67	0.44
9	0.47	0.29
10	0.65	0.51
11	0.82	0.63
12	0.69	0.44

The conclusion from Table 5.10 above is that, overall, pupils who are more able to associate words in WAT did better (higher facility value) than those pupils who are less able to associate words in WAT (lower facility value) in the multiple choice test. The reason is that the former group can associate the concepts strongly in their minds and so achieve better in the test related to these concepts, whereas the latter group have weak association of the concepts in their minds and so achieve less in the test related to these concepts.

The difference, in the facility value, between the two groups, was found statistically significant at 5% level for items **1, 8, 9, 11** and **12** when the method developed by Kellett (1978) was used to test the significance (Appendix 1.3).

5.7.4 The Effect of Introducing Distractors from WAT into MCQs:

Another major aim of this part of the research was to investigate the effect of introducing distractors, which appeared in pupils' responses in WAT, into the multiple choice test. In other words, how the response frequency of the option, which appeared in WAT, affects pupils' selection of that option. In order to examine this, the facility value of the key answer and the different distractors of each item was calculated and presented in Table 5.11 below.

Table 5.11: The facility value of the key answer and the different distractors of the biology multiple choice test

	Facility Value				
Item Number	**Key Answer**	**Distractor (1)**	**Distractor (2)**	**Distractor (3)**	**Omitting**
1	b) 0.47	a) 0.43	c) 0.01	d) 0.08	0.01
2	c) 0.88	a) 0.06	b) 0.03	d) 0.03	-
3	b) 0.41	a) 0.04	c) 0.11	d) 0.42	0.01
4	a) 1.00	b) -	c) -	d) -	-
5	a) 0.84	b) 0.01	c) -	d) 0.15	-
6	c) 0.68	a) 0.27	b) 0.02	d) 0.03	-
7	b) 0.93	a) 0.03	c) 0.01	d) 0.01	0.01
8	d) 0.57	a) 0.32	b) 0.06	c) 0.06	-
9	b) 0.39	a) 0.43	c) 0.10	d) 0.08	-
10	b) 0.60	a) 0.23	c) 0.02	d) 0.14	-
11	b) 0.73	a) 0.12	c) 0.10	d) 0.04	-
12	c) 0.58	a) 0.02	b) 0.24	d) 0.14	0.01

From the table above, the facility value of item four (example one of section 5.6.3.2) is 1.0. All pupils selected the option **a** (starch) which is the key answer in this item. At the same time this option has the biggest response frequency (21) in the WAT. Therefore, this item matches with our expectation that the option, which has high response frequency in WAT, is more likely to be selected by most of the pupils (high facility value).

Another example in which the option with a high response frequency in the WAT is more likely to be selected than the rest of the options is item eight of the test (example three of section 5.6.3.2). In this item, we predicted that option **d** (oxygen) will be chosen by most of the pupils as it has high response frequency in WAT. It was found that this prediction is correct. The facility value of this option is the largest compared to the rest of the options in the item (F.V of option **a** = 0.32, of option **b** = 0.06, of option **c** = 0.06 and of option **d** = 0.57).

The same explanation can be applied for items two, ten, and eleven. In these items, the correct option has both high facility value in the multiple choice test and biggest response frequencies in the WAT.

However, there are some exceptions, which for some reason, cannot be explained in the same manner. For example, in item five (example two of section 5.6.3.2), option d (plants) has the biggest response frequency (211) in the WAT but it was not a popular choice in the multiple choice test (F.V = 0.15). On the other hand, the key answer (sun) has only 47 response frequencies in the WAT but it has the facility value of 0.84. Most of the pupils selected the 'sun' as the main source of energy.

Another example of these exceptions is item seven. In this item, although option a (cacti) has lower response frequency in WAT compared to other options, it has a high facility value (0.93). The reason why 'cacti' did not appear strongly in the word association test is may because it can be found in desert which is different from the habitat where pupils live. However, when the pupils were directed in the stem of the question (item seven of the test) about the plant which can live for long time without rainfall, most of the pupils selected this option (option **a** = cacti) as the correct answer. The reasons for this may be as follows:

- It is the only name of the plants, given in the option list, live in a desert that can live for a long time without rainfall.
- This plant is given in the pupils' textbook as a typical example of the plants which can survive for long time without water.
- Simply, it is the key answer. In this case, it seems to be that the option being a key answer has an importance in pupils' selections.

The conclusion, which emerged from above, revealed the importance of the response frequencies of an option (obtained from the word association test) in pupils' selections to that option in the multiple choice test. However, it might be that there are other underlying factors which have also an importance in pupils' selections. Thus, much more research should be done in this area.

CHAPTER SIX: GENERAL DISCUSSION

CHAPTER SIX: GENERAL DISCUSSION

6.1 Introduction:

This chapter is going to discuss the results which were found in chapters three and five of the thesis. The researcher will give a general discussion about each part of the research.

6.2 Fixed Response: Multiple Choice Questions (MCQs)

6.2.1 Changing the Position of the Most Plausible Distractor (Strong Distractor):

The results of Chapter Three (section 3.2.2) showed that a small amount of editing can have a significant effect on the facility value of the MCQs. In this part of the research the editing was done by positioning the most plausible distractor next to, one position away or further away from the key answer.

The results revealed that placing the most plausible distractor in different positions to the key answer alters the responses to the key answer. The degree of difficulty of the item could be altered significantly. It has been found that in some cases the facility value of the item increases when the most plausible distractor (strong distractor) is located next to the key answer. However, it has been found that even placing the most plausible distractor away and far away from the key answer increases the facility value of the item.

Placing the most plausible distractor next to the key answer may help students to discriminate more effectively between them and therefore select the correct choice more easily than when they are away or further away from each other. On the other hand, when the most plausible distractor is located away or further away from the key answer this could make the comparison between the two more difficult (i.e. between the key answer and the most plausible distractor). Assuming that the other distractors are eliminated either physically or in the mind of the student.

The plan was to investigate the above factor (i.e. optical origin) practically at the university level using the computer based assessment package **'TRIADS'**. Unfortunately, we could not do it, as was mentioned in Chapter Three, section 3.2.5.4 because of the break down of the computer server.

However, if we assume that this investigation was successful, then we could make a prediction by saying that the results will show no significant difference in the facility value of the MCQs when the students were asked, firstly, to select the two definitely wrong options and secondly select the correct answer from the remaining options. By eliminating the two definitely wrong options, students are left with three options in the computer screen, which reduce the distraction from the other options. Consequently, students can make an effective comparison between the remaining options that leads more often to select the correct option.

Also it was found that not only the position of the most plausible distractor affects the facility value of the MCQs. There are may be several factors interacting to create this problem. One of these factors is unequal (uneven) length of the options in the question. It has been found that if all options of the item are of equal length, the facility value is higher when the most plausible distractor (strong distractor) is located next to the key answer. On the other hand, if the options are not of equal length the longer or the shorter option draws the responses from the key answer.

Finally what needs to be said, is that multiple choice test constructors should be aware of this fact that, especially when he selects the items from a pre – test, any alteration of the most plausible distractor position could have a significant effect on the facility value of the item (performance of the item). In other words, the pre – test statistics are no longer totally reliable for test construction purposes.

6.2.2 Changing the Position of the Key Answer:

The results of changing the position of the key answer (Chapter Three, section 3.2.4) showed that this can bring a significant difference in the facility value of the MCQs. It has been found that the facility value, in some items, can be change by 10% or more. Thus, it is not only the position of the most plausible distractor which affects the facility value of the MCQs but also the position of the key answer.

One possible explanation of why there is a difference in the facility values of the MCQs, when the key answer is altered, might just be related to the proximity of the key answer to the strong distractor. By changing the position of the key answer while keeping the position of the strong distractor still, can make for a clear close proximity of the key answer to the strong distractor. This may help students to discriminate between the two options more effectively and hence select the correct option.

This explanation was examined practically at the school level (Chapter Three, section 3.2.5). The researcher tried to design the test in such away that no other factors could affect the facility value except what is being investigated (i.e. the position of the key answer). In other words, the suggestions, which are reported in the literature about how to construct a good multiple choice questions, were taken into consideration. The results of this examination showed that the difference in the facility value of the MCQs when the key answer position was altered almost disappeared when pupils are asked, firstly, to eliminate the two definitely wrong options. After they had done this, pupils were left with only two options between which they could make an effective discrimination. As a result of this, pupils more often selected the correct option.

Moreover, as was stated in Chapter Three, section 3.2.6.3, we can look to the 'optical' factor above from the Information Processing Model points of view. The proximity effect upon facility values was almost eliminated when pupils/students were asked to tackle the questions in two steps mentioned above, which are:

- Eliminated the two options you think are definitely wrong.
- Now compare the remaining options and decide which is correct.

The elimination by physical means had the effect of removing the 'noise' of the question and simplifying the final choice. This had the effect of achieving stable facility values regardless of the initial relative position of the key answer and strong distractor. This strongly suggests that some 'optical' effect, or just a lowering of processing demand, leads to stable facility values.

Nevertheless, as was stated in Chapter Three, section 3.2.6.3, that this investigation was carried out with few items. Consequently, much more research should be done with larger number of items to confirm the findings of the current research.

In addition, it has been found that there is a tendency among students to favour certain positions of the key answer. The facility value of many items is higher when the key answer was placed in the top positions (**a**, **b** and **c**). It could be that students tend not to read every option in the item before selecting an answer (Fagley, 1987). Failure to read all options would result in a response bias for early options. Therefore, the facility value of those items in which key answer is placed in the top positions is more likely to be higher.

However, as stated in Chapter Two, section 2.2.1.1 that there are definitional differences among the educators about what evidence would demonstrate a positional response bias. In the current research, the researcher used very simple method to identify if there is a positional response bias (i.e. favour certain positions of the key answer). That is, the calculation was made across the five test versions to find out in how many items the facility value is higher when the key answer is placed in **a**, **b** and **c** positions and in how many items the facility value is higher when the key answer is placed in **d** and **e** positions. But is it really reflected if there is a positional response bias?

6.3 Fixed Response: Structural Communications Grids (SCGs)

Structural communication grids (SCGs) are another fixed response method of assessment which was put under the microscope for investigation in the current research. As was pointed out in Chapter Two, section 2.2.1.5, there was a lack of studies done in this method. Our research is just a starting point for further studies.

6.3.1 Changing the Positions of the Options within the Grid:

In the structural communication grids, change has been made to the positions of the options within the grid. The results showed that the facility value of the SCGs also suffers, to some extent, from changing the positions of the options within the grid.

This investigation was carried out at school and university levels. At the school level, the test was applied in two academic years. In general, the results showed that changing the positions of the options within the grid has an effect, to some extent, in the facility value of the questions in the two academic years. However, the number of sub – questions which showed a difference in the facility value in 1997/98 academic year is larger than the number of sub – questions which showed the difference in the facility value in 1998/99 academic year. But most of these sub – questions (1997/98 academic year) showed statistically no significant difference.

At the university level, the similar results were obtained. Also a small number of sub – questions showed a significant difference in the facility value when the positions of the options were rearranged within the grid.

There is no clear explanation why changing the positions of the options affects the facility value of the grid questions. It might be related to some psychological factors or just something optical. If it is something optical, then it could be that when the strong distractors are placed in the boxes next to the box(es), where the correct answer(s) are placed, helps students to discriminate more effectively between them. However. this is just an assumption and examining this practically is recommended the future research.

Another investigation was made to find out pupils'/students' performances in a test consisting of sub - questions from the two different versions of the test. It was applied to the school and university tests. The results showed that the difference in the average of the facility values of the test as whole, between the three occasions, is no more than 5%. This means that grid type questions showed a kind of stability in the average of the facility values which is very important to use as tool for pupils'/students' assessment.

There is another way to alter the positions of the options within the grid. It is by changing the shape of the grid. This was investigated in the current research but the sample was so small that we could not make a conclusion.

6.3.2 Two Methods of Scoring the Grid Questions:

In this investigation, two methods of scoring the grid questions were compared to find out if there is a difference in the pupils' rank order. The results revealed that there is no significant difference in pupils' ranking in the two methods. However, there are three points that should be concluded.

- In the Scottish Qualifications Authority scoring method, there are two types of questions (closed and open). For each type, different criteria have been used to allocate scores to the examinees. However, this is not the case in the method suggested by Egan, which treats all the questions in a similar way.

- In the Egan method, there is a possibility that a pupil is allocated credit for partial knowledge but in the Scottish Qualifications Authority scoring method no such provision exists in the closed type questions.

- The pupil is penalised for wrong selections in the method suggested by Egan. On the other hand, in the Scottish Qualifications Authority scoring method, a pupil is penalised only if he selects more or fewer than the required number despite the fact that one or two of his selection(s) were correct. This situation is applied for closed questions. In case of open questions, one mark is deducted from the pupil for each wrong selection.

The researcher's point of view recommended using the Egan method to score the grid questions because it is fairer and discriminates well between the pupils. Nonetheless. some teachers may refuse using the Egan method due to the fact it is difficult to do it by hand. This excuse is rejected because it can be done nowadays using the computer. which can score the questions very quickly.

6.4 Cognitive Styles and Assessment Methods:

Two cognitive styles were tested in this part of the research. These are field dependence – independence and convergence – divergence.

6.4.1 Field Dependent - Independent Cognitive Style and Assessment Methods:

A positive correlation emerged between pupils' performances in field dependent – independent test and their achievements in the grid questions (Chapter Five, section 5.3.1.1). We expected this correlation as was pointed out in Chapter Four, section 4.1.2.1 that field dependent – independent cognitive style related to overall academic achievement (Tinajero *et al.*, 1998). This indicates that pupils who scored highly in the grid questions tend to be field - independent and those who obtained low scores tend to be field - dependent.

Furthermore, it has been found that field - independent pupils performed better than field - dependent pupils in fixed response questions (in the grid form). It might be because field - independent pupils have better strategies and ways of handling the materials embedded in the question. In grid, both the signals and noise exist. It has been reported in the literature that field - independent pupils are more able than others to separate 'signal' from the 'noise'. Therefore, these pupils are using a considerable amount of their potential working space capacity in useful processing compared to other pupils who have difficulties in the separation between the 'signal' from the 'noise' and for whom the 'noise' captures some of their working space leaving a reduced space for useful processing (Al – Naeme, 1991).

There is one important point that we should mention here. In the current study, the number of boxes in each grid is six. The results showed that field - independent pupils perform better than field – dependent pupils in the fixed response questions (in the grid form). However, if we used grids with more than six boxes we expect the difference in the mean scores between the two groups will be higher than what we found in the research. The factor of number of boxes may affect the performance of each group of field dependent – independent cognitive style. A small number of boxes (as we have in the current research) means less 'noise' and therefore no discrimination between field – dependence and field – independence is likely. Conversely, the large number of boxes will operate against the field – dependent pupils, because the 'noise' level will increase and cause confusion for these pupils but have much less effect upon the field – independent pupils.

The same explanation could be applied why field - independent pupils performed better than field - dependent pupils in open response questions. Field - independent pupils have better strategies and ways of handling the materials embedded in the question, and are more able than others to separate 'signal' from the 'noise'.

Another small investigation was carried out to find out the performance of each group of field dependent – independent cognitive style in different versions of the grid questions. In other words, how is the performance of each group of field dependent – independent cognitive style affected when the options were altered within the grids? Unfortunately, the sample was so small that we could not draw a conclusion from it.

6.4.2 Convergent - Divergent Cognitive Style and Assessment Methods:

The results of the investigation of the effect of the convergent - divergent cognitive style on assessment methods showed that there is no correlation between pupils' performances in the convergent – divergent cognitive style test and their achievements in fixed response questions (multiple choice and grid questions).

Nevertheless, the reason why divergent pupils did not perform better than convergent pupils in the multiple choice questions (MCQs) could be that in MCQs converging is what is being asked for. The pupil knows he needs to look for only one correct answer. The divergers may see possibilities of correctness in more than one option and so make mistakes.

The results of the investigation of the performance of one group of convergent – divergent cognitive style in the grid questions showed that the mean score of divergent pupils is slightly over the mean score of convergent pupils. However, it has been found that this difference is statistically not significant.

Bahar (1999) found that there is a statistically significant difference between the mean score of convergent students and the mean score of divergent students in grid questions.

The divergent students achieved better than convergent students in grid questions at university level. However, the results of the current study did not agree with Bahar's results. In the current study, it was found that there was statistically no significant difference between the performance of convergent pupils and divergent pupils in the grid questions. The factor which seems likely to be the cause of this is related to the number of boxes in grid. In Bahar's (1999) study, grids with nine boxes (options) were used whereas in the current study the test consists of grids with six boxes (Appendix 4.1). It could to be that the grids with six boxes do not give enough possibilities (options) for the divergent pupils to show their superiority. Consequently, difference in performance between convergent and divergent pupils in the six boxes study, is not evident. On the other hand, the larger number should favour divergent pupils who may choose more boxes to provide a more complete answer.

6.5 The Relationship between Pupils' Knowledge Structures and Multiple Choice Test Structure:

The result of this part of the research showed that the option in the question, which is closer to the concept on which the stem of the question is based, is more likely to be selected than other options.

In Chapter Four, section 4.2.3.1, one model of how the knowledge is represented or stored in the memory has been discussed. In this model (Collins and Qullian (and Loftus) model), the knowledge is stored in memory in the form of an associative (semantic) network, with the nodes in this network corresponding to word concepts and the links to conceptual relations joining these concepts. This bit was used in the current research to construct a graphic map showing how eight chemical concepts are linked to each other in the pupils' minds on average (Figure 5.5, page 169).

Furthermore, the Collins and Qullian (and Loftus) model not only tells us how the knowledge is stored or represented in the pupils' minds but also describes how the information is accessed and retrieved from the network. The process of operating in the semantic memory that leads to retrieving the information is known as ***spreading – activation***. When the learner is asked a question about any concept, the activation spreads from node to node along the network links, making knowledge associated with particular sources of activation available for processing.

The above information was applied to our investigation. For example, the concept 'element' was found to be associated with other concepts in the pupils' minds (Figure 5.5). In the light of this, our multiple choice questions were designed to test the knowledge related to the concept '**element**'. When this concept is given in the stem of the question, the process of activation begins with the 'element' concept node. Then, it spreads throughout the network along all the connecting pathways until the target information is reached. The concepts, which are closer to the concept 'element', require less time for verification than those concepts farther away. The closer concepts are more likely to be selected as the correct answer(s) than the more distant concepts. This was tested practically by Collins and Quillian (1969) by what is called a ***sentence verification task*** (see Chapter Four, section 4.2.4 for more details).

The result of this investigation as was stated in Chapter Five, section 5.5.2 may not tell us everything about pupils' selections. However, it told us that the closer concepts (i.e. options) in the pupils' knowledge structures play apart in their selections in the multiple choice test than the more distant concepts (i.e. options). It might be that there are other factors, which we could not recognize in this investigation. Therefore, more research is needed to confirm this finding and explore the effects of other factors (if any) interacting with the factor of closeness.

6.6 Pupils' Performances in WAT and their Achievement in MCQs:

In this part of the research, the researcher was trying to find out the relationship between pupils' performances in a word association test (WAT) and their achievement in the multiple choice test dealing with the same concepts. In other words, how are pupils' performances in multiple choice test related to their ability to associate words in a word association test. The purpose of the word association test is different from its purpose of section 6.5. The purposes of the word association test in this part, as was listed in Chapter Five, section 5.6.1, are to:

- Map pupils' cognitive structures of eight biological concepts using WAT.
- Use WAT as an objective source for constructing MCQs in biology.
- Compare pupils' performance on WAT and their achievement on the multiple choice test.
- Investigate the effect of introducing, into the multiple choice test, distractors which appeared in pupils' responses in WAT.

The map of the eight biological concepts was produced. The map was used to identify which concepts are strongly associated and which are weakly associated in the pupils' minds. The map can be found in Chapter Five, section 5.7.1 (Figure 5.6, page 179).

Another investigation was made to compare the performance of those pupils who are more able to associate words in WAT and those pupils who are less able to associate words in WAT with their achievement in the multiple choice test. The results matched with the theory that the learner who can associate the concepts strongly in his mind achieves better in the test related to these concepts. On the other hand, the learner who has weak association of the concepts in his mind achieves less in the test related to these concepts.

Moreover, one major aim of this part of the research was to examine the effect of introducing distractors, which appeared in pupils' responses in WAT, into multiple choice test. The result revealed that an option, which has a high response frequency in WAT, is more likely to be chosen by pupils in the multiple choice question test. The opposite is true, the option, which has a low response frequency in WAT, is less likely to be selected in the test. Generally speaking, the effect of introducing distractors, which appeared in pupils' responses in WAT, into multiple choice test is that where the existing concepts are firmly linked in the cognitive structures facility values for questions testing these concepts are high.

However, as was pointed in Chapter Five, section 5.7.4, it might be that there are other hidden factors have an importance in pupils' selections in the multiple choice test. The current research was conducted to investigate the factor of response frequencies of the options, in word association test, in pupils' selections. Therefore, it is recommended that much more research should be done to reveal the importance of the other factors.

CHAPTER SEVEN: CONCLUSIONS, RECOMMENDATIONS AND SUGGESTIONS

CHAPTER SEVEN: CONCLUSIONS, RECOMMENDATIONS AND SUGGESTIONS

7.1 Conclusions of the Current Research:

We would like now to summarise the findings which have emerged from the present study.

7.1.1 The Major Findings:

a. The very important finding of this research highlights the problem of the small editing in the multiple choice questions. The results showed that changing the positions of the key answer and the most plausible distractor (i.e. changing the position of any option) has a significant effect in the facility value. It was found that the difference in the facility value of some items was 10% or more and most of these differences are statically significant. This finding will bring us back to the question of the validity of the multiple choice test. It seemed to be from the results that we are not only testing science but also testing other factors. This kind of positional response bias as Fagley (1987) argues would be an extraneous source of variance in test scores, reducing a test's content validity.

This problem plus the other problems listed in the literature (Chapter Two, section 2.2.1.1) make us skeptical about our assessment processes. Consequently, we should rethink our assessment based on multiple choice questions. If we want to conduct a good and fair assessment of our pupils/students, then we should rethink the use of multiple choice questions. Nevertheless, there are some advantages of multiple choice questions, but if we compare them with the disadvantages, we can outweigh the advantages if the disadvantages prevent us from conducting valid assessment.

There is no one overriding factor that we can be sure is the cause of the changes in the facility values of multiple choice questions when the positions of the key answer and the most plausible distractor are changed. However, one factor, which might be the cause, is something 'optical'. In other words, the position of the key answer relative to the strong distractor. It was found that when the most plausible distractor is located next to the key answer the facility value of the question tends to be higher than when it is away or further away. Close proximity may help students to discriminate more effectively between them and therefore more often select the correct choice.

This factor was examined practically at the school level. The results showed that when the pupils were asked first to remove the two options, which they think are definitely, wrong and then to select the correct one from the remaining two, the differences in the facility values were almost eliminated or at least strongly reduced.

However, this 'optical' factor can be explained by considering the Information Processing Model. This elimination before final choice procedure is just a lowering of processing demand. When the pupils were asked to remove the two definitely wrong options, they were actually removing the distracting 'noise' of the question which then simplified the final choice.

b. If we are going to attack the use of the multiple choice questions in our assessment, then we should offer an alternative method that can overcome or at least reduce the problems of these types of questions. As was mentioned in Chapter Two, section 2.2.1.5 structural communication grids provide an alternative fixed response method which can overcome most of the problems of multiple choice questions.

The results of the change made to the structural communication grids showed that changing the positions of the options within the grids also has, to some extent, an effect in the facility value. The difference in the facility value of some sub – questions is by 10% or more. These differences were found to be statically significant for some sub - questions. Nonetheless, it is not as serious as what has been found in multiple choice questions. The number of sub – questions (and questions) which has significant difference in the facility value is very small. This investigation was done at both school and university levels and similar results were obtained.

The above results and the advantages of grid question stated in Chapter Two, section 2.2.1.5 give a strong evidence that grid questions can be used instead of the multiple choice questions to assess students in a science curriculum. Most of the problems and limitations of multiple choice questions can be overcome or at least reduced when grid questions are used.

7.1.2 The Minor Findings:

a. In general, the facility value of an item is higher when the key answer is placed in top positions (**a**, **b** and **c**) than when it is placed in the bottom positions (**d** and **e**). In other words, students found the item easier when the key answer is placed in the upper positions.

b. It has been found that there is no significance difference in pupils' rank order when two scoring methods were used to score the grid questions; the Scottish Qualifications Authority scoring method and the method suggested by Egan. However, the method which is suggested by Egan seemed to be fairer and to discriminate better between pupils than the Scottish Qualifications Authority scoring method did. The advantages and disadvantages of each method were discussed in Chapter Two, section 2.2.1.5.

c. One of the important findings when checking the strategies pupils adopt to answer the grid questions, showed that most of the good pupils (top group) preferred the grid to be presented in **3** × **2** (the shape below).

A	B	C
D	E	F

Because:

- it was easy to compare between **THREE** options at the same time.
- pupils did not need to go up and down to search for the correct answer(s) as when the grid is presented in **2** × 3.
- the shape of the grid where pupils put their answers (shape below) is similar to the grid where the options were written. Consequently, this will help pupils to reduce the confusions and the mistakes that they may make while they are entering their answer(s).

A	B	C
D	E	F

On the other hand, most weak pupils (bottom group) preferred the grid to be presented in **2** ×**3** (the shape below):

A	D
B	E
C	F

Because it was easy to:

- compare between **TWO** options at the same time.
- see because it looks larger in size.

d. There is a positive correlation between pupils' performances in field dependent – independent cognitive style test and their achievement in the grid questions. The field – independent pupils had higher scores than field - dependent pupils in the grid questions.

e. The mean score of field – independent pupils is higher than the mean score of field - dependent pupils in both fixed and open response questions. The fixed response questions were presented to pupils in the grid form.

f. There is no correlation between pupils' performances in convergent – divergent cognitive style test and their achievement in fixed response questions.

g. The mean score of divergent pupils is greater than the mean score of convergent pupils in multiple choice questions and structural communication grids. However, the difference in the mean scores between the two groups in both types of questions was found not to be statically significant.

h. It was found that the option in the question, which is closer in the graphic map to the concept on which the stem of the question is based, is more likely to be selected than other options. This was tested practically at the school level and the results supported this hypothesis (Chapter Five, section 5.5).

i. There is a positive correlation between pupils' performances in a word association test (WAT) and their achievement in the multiple choice test dealing with the same concepts.

j. Pupils who are more able to associate concepts in WAT perform better than those who are less able to associate concepts in WAT in the multiple choice test. The facility value of the MCQs for those pupils who are more able to associate concepts in WAT is larger than the facility value of those pupils who are less able to associate concepts in WAT. The difference in the facility value between the two groups was found to be statistically significant for most questions.

k. The effect of introducing, into a multiple choice test, distractors which appeared in pupils' responses in WAT revealed that in most cases the option which has high response frequencies in WAT is most likely to be selected among other options in the multiple choice questions.

7.2 Suggestions for Teachers and Lecturers:

The researcher believes that any research in science education should deliver a message to educators or teachers from its findings and should contribute to educational development as whole, otherwise it is a waste of time. The present research conveys an educational message to teachers and lecturers who are responsible for assessing learners about how to conduct better assessment at the school and university levels. The following are the recommendations and suggestions which are derived from the present study.

a. Grid questions should take an important place in the assessment process. Although we have seen that changing the positions of the options within a grid cause, to some extent, a difference in the facility value of the grid questions. This also happen in the multiple choice questions (MCQs) to a greater extent. Grid questions have many advantages that are not found in the MCQs. The following are some of the advantages of grid questions.

- They can be used as an alternative tool to the traditional methods of diagnostic assessment of the pupil's prior knowledge of specific concepts in science.
- They allow an objective measure of the higher level intellectual skills.
- They are a good way of testing conceptual understanding.
- They are objectively markable with a highly reliable computer technique.
- They can be used to test some of the skills used in essay type questions.
- The contents of the boxes can be varied, so that they can be made suitable for visual as well as verbal thinkers.
- There is less opportunity for guessing compared to multiple choice questions.
- Teacher can ask many questions based on one grid.
- They can go beyond asking selection questions only. Teachers can ask students to put their selections in a logical order. This helps teachers to find out how the ideas are linked together in the pupil's/student's mind and to discover why a student makes a mistake.

However, to improve the effectiveness of the grid questions, the following should be taken into consideration while constructing and marking the questions:

- The options in the grid should be used as correct answers in more than one question. This will help the assessor to eliminate or at least reduce guessing.
- The assessor should not announce in the stem of the question how many boxes the student/pupil should select to answer the question. In other words, the open questions in the Scottish Qualifications Authority scoring method should be used in the grid.
- Using the appropriate method of scoring the selection and the logical order such as the method suggested by Egan (**Coefficient of Confusion**) to score the selection and the Mackenzie's third method (**Before and Next to**) to score the logical order.
- The number of answers should not exceed half (1/2) of total number of boxes available in the grid; otherwise students/pupils might be given unfair marks.
- The number of boxes in each grid should be limited. In other words, the number of boxes should match the level of development of the pupils, and at the same time be more than in a normal multiple choice test. This will help the assessor firstly, to allow divergent pupils room to show their imagination and secondly to eliminate guessing.

b. Having exposed serious limitations in the validity of multiple choice questions, the researcher would urge caution in their use. He would strongly recommend that they should not be used as the sole method of assessment, but only as one component in a test using a variety of assessment methods. Having done this, the examiners should take into consideration the suggestions, which are reported in the literature, how to design good multiple choice questions.

c. Different methods of assessment should be used to match the different pupils'/students' cognitive styles. This is due to the fact that it is difficult to find one type of assessment method that is suitable for all kinds of the cognitive styles. For example, divergent pupils may perform better in grid questions, with more than six boxes, than in multiple choice questions. On the other hand, convergent pupils may perform better in multiple choice questions than in grid questions.

d. Applying word association techniques and drawing the graphic map is hard work. Nonetheless, it is a valuable technique to help examiners to construct good multiple choice questions. It can be used in different ways:

- The graphic map as an objective source of choosing the options from different positions of the pupils' knowledge structures.
- The response which are given by pupils as an objective source of selecting options to include in a multiple choice format.
- As a diagnostic tool by which the teacher can identify the misconceptions that his pupils/students have in a particular area of the curriculum before the exam takes place.

7.3 Suggestions for Future Work:

This research was attempting to find out answers to some questions related to some assessment methods, mainly fixed response questions. The answers have been found to some of these questions, but these answers do not give the complete picture. Furthermore, new questions have arisen during the course of research study. The following are some suggestions for future research.

a. The new computer program "Tripartite Interactive Assessment Delivery System **(TRIADS)**" can be used to examine if the differences in the facility values of the MCQs when the positions of the key answer and the most plausible distractor have been altered is due to an 'optical' effect. This investigation can be done in different science subjects such as biology, chemistry and geology.

b. It is recommended that much more research should be done to find out the effect of changing the shape of the grid (i.e. another way of changing the positions of the options).

c. A further research study may be necessary to examine more factors affecting pupils'/students' performances in grid questions. For example, it would be interesting to find out the relationship between the optimum number of boxes in the grid and pupils'/students' working memory space.

d. Related to field dependent – independent cognitive style, it is recommended for future work to look at how the performances of low capacity and high capacity field – dependent and low and high capacity field – independent in fixed and open response questions.

e. It could be very useful to find out how the performance of divergent pupils differ from convergent pupils in grid questions of more than six boxes. This can be done in different fields of science (biology, chemistry, physics and geology).

f. It could be useful to investigate the effect of choosing the options of multiple choice test from different positions in the pupils' knowledge structures in other fields of science such as biology and physics to support or reject the finding of the current research.

REFERNCES

REFERENCES

1. Abdullah, A. and Scaife, J. (1997) "Using interviews to assess children's understanding of science concepts", *School Science Review*, **78**(285), 79- 83.

2. Ace, M. and Dawis, R. (1973) "Item structure as a determinant of item difficulty in verbal analogies", *Educational and Psychological Measurement*, **33**, 143-149.

3. Adams, K., Ginn, B. and Ruddick, D. (1998) "Independent learning in an introductory module in biological chemistry: use of Question Mark ™ software to provide an assessment tool and tutorial support, **2**(2), 40-44.

4. Aiken, L. (1997) *Psychological Testing and Assessment*, Boston, Allyn and Bacon.

5. Alamolhodaei, H. (1996) "A study in higher education calculus and students' learning styles", *Ph.D. Thesis*, Glasgow, University of Glasgow.

6. Al - Naeme, F. (1989) "A study of some psychological factors affecting performance in chemistry at secondary and tertiary levels", *MSc. Thesis*, Glasgow, University of Glasgow.

7. Al - Naeme, F. (1991) "The influence of various learning styles on practical problem-solving in chemistry in Scottish secondary schools", *Ph.D. Thesis*, Glasgow, University of Glasgow.

8. Amir, R. and Tamit, P. (1995) "Proposition generating task (PGT): a measure of meaningful learning and of conceptual change", *Journal of Biological Education*, **29** (2), 111-118.

9. Anderson, J. and Pirolli, P. (1984) "Spread of Activation", *Journal of Experimental Psychology: Learning, Memory and Cognition*, **10**(4), 791-798.

10. Ashcraft, M. (1976) "Priming and property dominance effects in semantic memory", *Memory and Cognition*, **4**(5), 490-500.

11. Ashcraft, M. (1994) *Human Memory and Cognition* (2nd. ed.), New York, Harper Collins College Publishers.

12. Ausubel, D. (1963) *The Psychology of Meaningful Verbal Learning*, New York, Grune and Stratton.

13. Bahar, M (1999) "Investigation of biology student's cognitive structure through word association tests, mind maps and structural communication grids", *Ph. D Thesis*, Glasgow, University of Glasgow.

14. Bahar, M., Johnstone, A. and Sutcliffe, R (1999) "Investigation of student's cognitive structure in elementary genetics through word association tests", *Journal of Biological Education*, **33**(3), 134 - 141.

15. Bailey, P. (1999) "Assessment of chemistry degrees", University Chemistry Education, **3**(2), 64 - 67.

16. Balla, J. and Boyle, P. (1994) "Assessment of student performance: a framework for improving practice", *Assessment and Evaluation in Higher Education*, **19**(1), 17-28.

17. Bannister, R., Cooper, B. and Foy, J. M. (1967) "A note on the use of objective tests in physiology", *Journal of Biological Education*, **1**(1), 47-56

18. Batting, W. and Montague W. (1969) "Category norms for verbal items in 56 categories: A replication and extension of the connecticut category norms", *Journal of Experimental Psychology Monograph*, **80**(3), Pt. 2, 1-46.

19. Baummann, A., Bloomfield, A. and Linda, R. (1997) *Becoming a Secondary School Teacher*, London, Hodder and Stoughton.

20. Billington, H. L. (1997) "Poster presentations and peer assessment: novel forms of evaluation", *Journal of Biological Education*, **31**(3), 218-220.

21. Black, P. J. (1993) "Formative and summative assessment by teachers", *Studies in Education*, **21**, 49-97.

22. Black, P. (1998) *Testing: Friend or Foe? Theory and Practice of Assessment and Testing*, London, The Falmer Press.

23. Blum, A, and Azencot, M. (1986) "Multiple - choice versus equivalence essay questions in a national examination", *European Journal of Science Education*, **8**(2), 225-228.

24. Blunch, N. J. (1984) "Position bias in multiple - choice questions", *Journal of Marketing Research*, **21**, 216-220.

25. Bowden, E. M. (1985) "Accessing relevant information during problem solving: time constraints on search in the problem space", *Memory and Cognition*, **13**(3), 280 - 286.

26. Bowen, C. W. (1997) "Item design considerations for computer - based testing for student learning in chemistry", *Manuscript Paper No. 19970185.*

27. Bray, E (1986) "Written responses: essays, guided responses, structure questions" in Lloyd - Jones, R., Bray, E., Johnson, G. and Currie, R. (eds.) *Assessment from Principle to Action*, London, Macmlion Education Ltd., 55-69.

28. Bresnock, A. E., Graves, P. E. and White, N. (1989) "Multiple - choice testing: questions and response position", *Journal of Economic Education*, **20**(3), 239-245.

29. Brodzinsky, D. M. (1982) "Relationship between cognitive style and cognitive development: a 2 - years longitudinal study", *Development Psychology*, **18**, 617 - 626.

30. Brooks, V. (1993) "The resurgence of external examining in Britain: a historical review", *British Journal of Educational Studies*, **41**(1), 59-72.

31. Buchan, A. S. and Jenkins, E. W. (1992) "The internal assessment of practical skills in science in England and Wales, 1960-1991: some issues in historical perspective", *International Journal of Science Education*, 14(4), 367-380.

32. Burton, L. and Haines, C. (1997) "Innovative in teaching and assessing mathematics at university level", *Teaching in Higher Education*, 2(3), 273-293.

33. Burton, R. and Miller, D. (2000) "Why tests are not as simple as a, b or c.", *The Times Higher* (4 Feb. 2000).

34. Cachapuz, A and Maskill R. (1987) "Detecting changes with learning in the organization of knowledge: use of word association tests to follow the learning of collision theory", *International Journal of Science Education*, 9(4), 491-504.

35. Cachapuz, A and Maskill R. (1989) "Using word association in formative classroom tests: following the learning of Le Chatelier's principle", *International Journal of Science Education*, 11(2), 235-246.

36. Campbell, N. (1996) *Biology* (4th. ed.) California, The Benjamin/Cumming Publishing Company, Inc.

37. Capper, J. (1996) *Testing to Learn - Learning to Test*, Washington DC, International Reading Association and the academy for Educational Development.

38. Carrie, J. (1984) "The growth of the concept of energy - A longitudinal study in Scottish secondary schools", *Ph. D Thesis*, Glasgow, University of Glasgow.

39. Cassels, A. (1991) *Remembering and Forgetting*, Leicester, The British Psychological Society.

40. Cassels, J. R. T., and Johnstone, A. H. (1984) "The effect of language on student performance on multiple choice tests in chemistry", *Journal of Chemical Education*, **61**(7), 613-515.

41. Cassels, J. R. T., and Johnstone, A. H. (1985) "Do you still beat your wife? and other interesting questions", *CHEMTECH, DECEMBER ISSUE*, 716-719.

42. Celegg, F. (1990) *Simple Statistics*, Cambridge, Cambridge University Press.

43. Chang, T. (1986) "Semantic memory: facts and models", *Psychological Bulletin*, **99**(2), 199-220.

44. Chapman, B. R. (1973) "Assessment and science education", *School Science Review*, **55**(191), 380-385.

45. Child, D. (1997) *Psychology and the Teacher* (6th ed.) London, Cassell.

46. Chmielewski, T., Danserfau, D. and Moreland, J. (1998) "Using common region in node-link displays: The role of field dependence/independence", *The Journal of Experimental Education*, **66**(3), 197-207

47. Cizek, G. J. (1994) "The effect of altering the position of options in a multiple choice examination", *Educational and Psychological Measurement*, **54**(1), 8-20.

48. Collins, H., Johansen, J. and Johnson, J. (1976) *Educational Measurement and Evaluation (2nd. ed.)*, Illinois, Scott, Foresman and Company.

49. Collins, A. and Loftus, E. (1975) "A spreading - activation theory of semantic process", *Psychological Review*, **82**(6), 407-428.

50. Collins, A. and Quillian, M. (1969) "Retrieval time for semantic memory", *Journal of Verbal Learning and Verbal Behavior*, **8**, 432-438.

51. Conrad, C. (1972) "Cognitive economy in semantic memory", *Journal of Experimental Psychology*, **92**(2), 149-154.

52. Cronbach, L. (1950) "Further evidence on response sets and test design", *Educational and Psychological Measurement*, **10**, 3-31.

53. Currie, R. (1986a) "Objective testing" in Lloyd - Jones, R., Bray, E., Johnson, G. and Currie, R. (eds.) *Assessment from Principle to Action*, London, Macmlion Education Ltd., 47-54.

54. Currie, R. (1986b) "Examinations in schools" in Lloyd - Jones, R., Bray, E., Johnson, G. and Currie, R. (eds.) *Assessment from Principle to Action*, London. Macmlion Education Ltd., 119-137.

55. Daniels, D. J. (1968) "Objective testing", *School Science Review*, **49**(170), 19-25.

56. Davis, J and Frank, B. (1979) "Learning and memory of field independence - dependence individuals", *Journal of Research in Personality*, **13**, 469-479.

57. Davis, J. C. and Mausert, P. (1981) "Effective testing and evaluation in chemistry", in Lippincoot W. and Heikkinen, H. (eds.) *Source Book for Chemistry Teachers*, American Chemical Society.

58. Deese, J. (1962) "On the structure of associative meaning", *Psychological Review*, **69**(3), 161 -174.

59. Dempster, J. A. (1994) "Question Mark designer for windows" *Active Learning* (1), 47-50.

60. Dixon, R. A. (1994) "Evaluating and improving multiple choice papers: true - false questions in public health medicine", *Medical Education*, **28**, 400-408.

61. Doherty, M. (1998) "Teaching and learning technology concepts in secondary schools: a study of psychological factors affecting pupil performance", *MSc. Thesis*, Glasgow, University of Glasgow.

62. Duncan, I. (1974) "The use of programmed learning materials to investigate learning processes in difficult areas in school chemistry", *MSc. Thesis*, Glasgow, University of Glasgow.

63. Ebbing, D. (1996) *General Chemistry*, Boston, Houghton Mifflin Company.

64. Ebel, R. (1972) *Essentials of Educational Measurement*, New Jersey, Prentice - Hall,Inc.

65. Egan, K. (1972) "Structural communication: a new contribution to pedagogy". *Programmed Learning Educational Technology*, 63-78.

66. Egan, K. (1973) "Measuring the ability to organise knowledge", *Educational Technology*, 63-78.

67. Egan, K. (1976) *Structural Communication*, California, Fearon Publishers.

68. El - Banna, H. (1987) "The development of a predictive theory of science education based upon information processing hypothesis". *Ph.D. Thesis*, Glasgow, University of Glasgow.

69. Entwistle, N. J. and Entwistle, A. (1991) "Constructing forms of understanding for degree examinations: the student experience and its implications", *Higher Education*, **22**, 205-227.

70. Entwistle, N., Thompson, S. and Tait, H. (1992) "Guidelines for promoting effective learning in higher education" Edinburgh, University of Edinburgh.

71. Fagley, N. S. (1987) "Positional response bias in multiple - choice tests of learning: its relation to testwiseness and guessing strategy", *Journal of Educational Psychology*, **79**(1), 95-97.

72. Fairbrother, B. (1993) "Problems in the assessment of scientific skills", in Edwards, D., Scanlon, E. and west, D. (eds.), *Teaching, Learning and Assessment in Science Education*, London, Paul Champman Publishing Ltd., 237-249.

73. Fenderson, B. A., Damjanov, I., Robeson, M. R., Veloski, J. J and Rubin, E. (1997) "The virtues of extended matching and uncued tests as alternatives to MCQs", *Human Pathology*, **28**(5), 526-532

74. Frederiksen, N., Ward, W. and Carlson, B. (1980) "Can multiple choice tests measure creativity?", *FINDINGS, ETS Research in Postsecondary Education*. **6**(1).

75. Friedman, S. and Cook, G. (1995) "Is an examinee's cognitive style related to the impact of answer changing on multiple - choice tests?", *Journal of Experimental Education*, **63**(3), 199-213.

76. Friel, S. (1976) "An investigation into objectives testing in science in a Scottish secondary school", *MSc. Thesis*, Glasgow, University of Glasgow.

77. Friel, S. and Johnstone, A. H. (1978) "Scoring systems which allow for partial knowledge", *Journal of Chemical Education*, **11**(11), 717-719.

78. Friel, S. and Johnstone, A. H. (1979) "Does the position of the answer in a multiple - choice test matter", *Education in Chemistry*, **16**, 175.

79. Friel, S. and Johnstone, A. H. (1988) "Making test scores yield more information", *Education in Chemistry*, **25**(2), 46-49.

80. Gallacher, D. (1984) "An investigation into the use of structural communication assessment units in diagnostic testing in science", *M.Ed. Thesis*, Glasgow, University of Glasgow.

81. Gamble, R. (1987) "An assessment paradox", *School Science Review*, **68**(244), 554-556.

82. Garskof, B. and Houston, J. (1963) "Measurement of verbal relatedness: An idiographic approach", *Psychological Review*, **70**(3), 277-288.

83. Gibbs, G. (1995) *Assessing Student Centred Courses*, Oxford, The Oxford Centre for Staff Development.

84. Giddings, G. Hofstein, A. and Lunetta, V. (1991) "Assessment and evaluation in science laboratory" in Woolnough, B. (ed.), *Practical Science*, Buckingham, Open University Press, 167-177.

85. Gott, R. and Duggan. S. (1995) *Investigative Work in Science Curriculum*, Buckingham, Open University Press.

86. Gray, C. (1997) "A study of factors affecting a curriculum innovation in university chemistry", *Ph.D. Thesis*, Glasgow, University of Glasgow.

87. Gronlund, N. (1968) *Constructing Achievement Tests*, New Jersey, Prentice - Hall, INC.

88. Gronlund, N. (1971) *Measurement and Evaluation in Teaching* (2nd ed.), New York, The Macmillan Company.

89. Guilford, J (1980) "Cognitive styles: what are they?", *Educational and Psychological Measurement*, **40**, 715-735.

90. Hager, P., Gonczi, A. and Athanasou, J. (1994) "General issues about assessment competence" *Assessment and Evaluation in Higher Education*, **19**(1), 3-16.

91. Hall, C. (1988) *Analysing and Combining Marks*, Wellington, Authority for Advanced Vocational Awards.

92. Handy, J. and Johnstone, A. (1973a) "How students reason in objective tests", *Education in Chemistry*, **10**, 99-100.

93. Handy, J. and Johnstone, A. (1973b) "Reproducibility in objective testing", *Education in Chemistry*, **10**(2), 47-48.

94. Hasson, F. (1988) "A study of psychological factors affecting student performance in biology", *MSc. Thesis*, Glasgow, University of Glasgow.

95. Head, J. J. (1965) "The construction of O' level examination questions in biology", *School Science Review*, **46**(160), 634-645.

96. Henson, R. C. and Stumbles, A. M. (1979) "Modern mathematics topics in chemistry", *The School Science Review*, 60(212), 446-458.

97. Heywood, J. (1977) "Examining in second- level education", Dublin, Association of Secondary Teachers.

98. Hodson, D. (1986) "A checklist of questions for science curriculum evaluation", *School Science Review*, **68**(243), 328-334.

99. Hudson, B. (1969) "Guessing in multiple - choice tests", *The School Science Review*, **LI**(175), 437-439.

100. Hudson, L. (1962) *Contrary Imaginations*, London, Penguin Books.

101. Hurd, P. D. (1969) *New Directions in Teaching Secondary School Science*, Chicago, Rand McNally and Company.

102. Jackson, B. (1988) "A comparison between computer - based and traditional assessments tests and their effects on pupil learning and scoring", *School Science Review*, **69**(249), 809-815.

103. Jeans, A. and Jeans, S. (1978) "Interactive assessment", *School Science Review*, **60**(211), 359-361.

104. Johnsone, E. (1977) "The use of modern mathematics in chemistry teaching", *The School Science Review*, **59**(206), 73-75.

105. Johnsone, P. (1964) "Associative meaning of concepts in physics", *Journal of Educational Psychology*, **55**(20), 84-88.

106. Johnsone, P. (1965) "Some psychological aspects of subject - matter structure", *Journal of Educational Psychology*, **58**(2), 75-83.

107. Johnsone, P. (1967) "Word relatedness and problem - solving in high - school physics", *Journal of Educational Psychology*, 56(4), 217-224.

108. Johnsone, P. (1969) "On the communication of concepts in science", *Journal of Educational Psychology*, **60**(1), 32-40.

109. Johnstone, A. (1988) *Meaning Beyond Readability*, Surrey, Southern Examining Group.

110. Johnstone, A. (2000a) "Teaching of chemistry – logical or psychological?", *Chemistry Education: Research and Practice in Europe,* **1**(1), 9-15.

111. Johnstone, A. (2000b) "Chemical education research: Where from here?", University Chemistry Education, **4**(1), 32-36.

112. Johnstone, A. and Ambusaidi, A. (2000) "Fixed response: what are we testing?", *Accepted for publication in Chemistry Education: Research and Practice in Europe,* **1**(3).

113. Johnstone, A. and Al-Naeme, F. (1991) "Room for scientific thought", *International Journal of Science Education*, **13**(2), 187-192.

114. Johnstone, A., Bahar, M. and Hansell, M. (2000) "Structural communication grids: a valuable assessment and diagnostic tool for science teachers", *Journal of Biological Education*, **34**(2), 87-89.

115. Johnstone, A. and El – Banna, H. (1989) "Understanding learning difficulties – a predictive research model", *Studies in Higher Education*, **14**(2), 159-168.

116. Johnstone, A.; MacGuire, P.; Friel, S. and Morrision, E.W. (1983) "Criterion - referenced testing in science - thoughts, worries and suggestions", *School Science review*, **64**(229), 626-634.

117. Johnstone, A.; McAlpine, E. and MacGuire, P. (1986) "Branching trees and diagnostic testing", *Journal of Further and Higher Education in Scotland,* **11**, 4-7.

118. Johnstone, A. and Moynihan, A. (1985) "The relationship between performance in word association tests and achievement in chemistry", *The European Journal of Science Education*, 7(1), 57-66.

119. Johnstone, A. H. and Mughol, A. R. (1979) "Testing for understanding", *School Science Review*, **61**, 147-150.

120. Johnstone, A. and Sharp, DWA. (1972) "Objective testing in tertiary education chemistry courses", *Chemistry in Britain*, **8**(2), 66-68.

121. Kellett, N. (1978) "Studies on the perception of organic chemical structures", *Ph.D. Thesis*, Glasgow, University of Glasgow.

122. Kellington, S. A. and Mitchell, A. C. (1980) "Designing an assessment system on the principles of criterion - referenced measurement", *School Science Review*, **61**(217), 765-770.

123. Kempa, R and Nicholls, C (1983) "Problem - solving ability and cognitive structure- an exploratory investigation", *European Journal of Science Education*, **5** (2), 171-184.

124. Kempa, R and Hodgson, G. (1976) "Levels of concept acquisition and concept maturation in students of chemistry", *British Journal of Educational Psychology*, **46**, 253-260.

125. Knutton, S. (1994) "Assessing children's learning in science" in Wellington, J., Henderson, J., Lally, J. Knutton, S. and Nott, M. (eds.), *Secondary Science Contemporary Issues and Practical Approaches*, London, Roultedge.

126. Leonard, L., Macteer, E. and Wilson, R. (1996) "Report of objective testing study group", Glasgow, Institute of Biomedical and Life Sciences, University of Glasgow.

127. Leonard, L., Sutcliffe, R. and MacAteer, E. (1997) "Development and evaluation of computer based assessment for first year biology students", *Paper presented at ALT - C97*, University of Wolverhampton.

128. Leonard, L. and Sutcliffe, R. (1999) "Computer - based assessment using TRIADS: the Glasgow experience", *Life Sciences Educational Computing*, **10**(3), 17-19.

129. Lloyd - Bostock, S. (1979) "Convergent - divergent thinking and arts - science orientation", *British Journal of Psychology*, **70**, 155-163.

130. Lloyd - Jones, R.(1986) "An overview of assessment" in Lloyd - Jones, R., Bray, E., Johnson, G. and Currie, R. (eds.) *Assessment from Principle to Action*, London, Macmlion Education Ltd., 1-13.

131. Lorsbach, A. W., Tobin, K., Briscove, C. and Ulerick, S. (1992) "An interpretation of assessment methods in middle school science", *International Journal of Science Education*, **14**(3), 305-317.

132. MacGuire, P. R. P. and Johnstone, A. H. (1987) "Techniques for investigating the understanding of concepts in science", *International Journal of Science Education*, **9**(5), 565-577.

133. Macintosh, H. and Morrison, R. (1969) *Objective Testing*, London, University of London Press LTD

134. Mackenzie, D (1997) *TRIAD: A Computer Based Assessment Software*, University of Derby Web Site.
http://www.derby.ac.uk/assess/newdemo/mainmenu.html

135. Mackenzie, D (1999) 'Recent developments in the Tripartite Interactive Assessment Delivery System (TRIADS)', *The paper has been updated from the original publish in the Proceedings of the Third International Computer Assessment Conferences*, Loughborough, University of Loughborough.

136. Marcus, A. (1963) "The effect of correct response location on the difficulty level of multiple - choice questions", *Journal of Applied Psychology*, 47(1), 48-51.

137. Marshall, J. and Hales, L. (1971) *Classroom Test Construction*, Reading. Addison-Wesley Publishing Company.

138. Mavrommaties, Y. (1997) "Understanding assessment in the classroom: phases of the assessment process - the assessment episode", *Assessment in Education*, 4(3), 381-399.

139. McGrath, M. and Ingham, A. (1992) "Using art activities in the assessment of science in the primary school", *School Science Review*, **73**(264), 33-47.

140. McNamara, W. and Weitzman, E. (1945) "The effect of choice placement on the difficulty of multiple - choice questions", *The Journal of Educational Psychology*, **36**, 103-113.

141. Moynihan, T. (1981) "Word association techniques for the study of cognitive structure of chemical concepts in Scottish secondary schools", *M.Sc. Thesis*, Glasgow, The University of Glasgow.

142. Nahinsky, I, Morgan, M and Oeschger, D. (1979) "Cognitive strategies, field dependence, and the abstraction process", *Journal of Research in Personality*, 13, 490-504.

143. Ndalichako, J. and Rogers, W. T. (1997) "Comparison of Finite State Score Theory, Classical Test Theory, and Item Response Theory in scoring multiple - choice items", *Educational and Psychological Measurement*, **57**(4), 580-589.

144. Nitko, A. J. (1983) *Educational Tests and Measurement: An Introduction*, New York, Harcourt Brace Jovanovich, INC.

145. Noble, J. and Frank, B. (1985) "Field - independence/dependence and verbal restructuring", *The Journal of Experimental Education*, 54(1), 28-33.

146. Pamphett, R. and Farnill, D. (1995) "Effect of anxiety on performance in multiple choice examination", *Medical Education*, **29**, 297-302.

147. Payne, S. (1951) *The Art of Asking Questions*, Princeton, Princeton University Press.

148. Preece, P. (1976) "Mapping cognitive structure: A comparison of methods", *Journal of Educational Psychology*, **68**(1), 1-8.

149. Preece, P. (1978) "Exploration of semantic space: review of research on the organization of scientific concepts in semantic memory", *Science Education*. **62**(4), 547-562.

150. Question Mark Computing (1996) *Question Designer: Testing and Assessment for Windows*, London, Question Mark Computing Ltd.

151. Race, P. (1997) Changing Assessment to Improve Chemistry Learning, A document founded by Project Improve through the *hefce* Fund for the Development of Teaching and Learning.

152. Rasp, S. L. (1998) "Toward more performance evaluation in chemistry", *Journal of Chemical Education*, **75**(1), 64-66.

153. Reisberg, D. (1997) *Cognition: Exploring the Science of the Mind*, New York, W. W. Norton and Company, Inc.

154. Renfrew, R. and Conquest, N. (1995) Standard Grade Chemistry with Answers, London, Hodder and Stoughton.

155. Rickards, J., Fajen, B., Sullivan, J. and Gillespie, G. (1997) "Signalling, notetaking, and field independence - dependence in text comprehension and recall", *Journal of Educational Psychology*, **89**(3), 508-517.

156. Rips, L., Shoben, E. and Smith, E. (1973) "Semantic distance and the verification of semantic relations", *Journal of Verbal Learning and Verbal Behavior*, **12**, 1-20.

157. Robson, C. (1994) *Experiment, Design and Statistics in Psychology (3ed ed.)*. Middlesex, Penguin Books.

158. Roediger, H. and Neely, J. (1982) "Retrieval blocks in episodic and semantic memory", *Canadian Journal of Psychology*, **36**, 213-242.

159. Rowntree, D. (1987) *Assessing Students: How Shall we Know them?*, London, Kogan Page.

160. Runco, M. (1986) "Divergent thinking and creative performance in gifted and nongifted children", *Educational and Psychological Measurement*, **46**, 375-383.

161. Russell, J (1992) *General Chemistry* (2nd. ed.), New York, McGraw-Hill, Inc.

162. Schawirth, LWT, Van Der Vleuten, CPM and Donkers, HHLM (1996) "A closer look at cueing effects in multiple - choice questions", *Medical Education*, **30**, 44-49.

163. Schawirth, LWT, Van Der Vleuten, CPM, Stoffers, HE JH and Peperkamp, AGW (1996) "Computerized long - menu questions as an alternative to open - ended questions in computerized assessment", *Medical Education*, **30**, 50-55

164. Schmidt, H. J. and Beine, M. (1992) "Setting multiple choice - tests", *Education in Chemistry*, **29**(1), 19-21.

165. Scottish Certificate of Education (1997) *Higher Grade Chemistry*, Glasgow, Robert Gibson-Publisher.

166. Scottish Certificate of Education (1997) *Standard Grade Chemistry*, Glasgow, Robert Gibson-Publisher.

167. Scouller, K. (1998) 'The influence of assessment methods on students' learning approaches: Multiple choice question examination versus assignment essay', *Higher Education,* **35**, 453-472.

168. Shaw, K. E. (1993) "Moderation in assessment: the case of MAP-SW", *Research Papers in Education*, **8**(2), 227-249.

169. Shavelson, R. (1972) "Some aspects of the correspondence between content structure and cognitive structure in physics instruction", *Journal of Educational Psychology,* **63**(3), 225-234.

170. Shavelson, R. (1973) "Learning from physics instruction", *Journal of Research in Science Teaching*, **10**(2), 101-111.

171. Shavelson, R. (1974) "Methods for examining representations of a subject - matter structure in a student's memory", *Journal of Research in Science Teaching*, **11**(3), 231-249.

172. Simpson, M. (1993) "Diagnostic assessment and its contribution to pupils' learning" in Edwards, D., Scanlon, E. and west, D. (eds.), *Teaching, Learning and Assessment in Science Education*, London, Paul Champman Publishing Ltd., 225-236.

173. Skevington, J. (1994) "Assessing and evaluation in science education" in Levinson, Ralph (ed.) *Teaching Science*, London, Roultedge, 115-128.

174. Smith, E., Shoben, E. and Rips, L. (1974) "Structure and process in semantic memory: a featural model for semantic decisions", *Psychological Review,* **81**, 214-241.

175. Staff Development and Educational Methods Unit, "Selecting and planning methods of assessment", Manchester, Manchester Polytechnic, 19-33.

176. Stenhouse, D. (1969) "Examination selection pressures in biology", *Journal of Biological Education*, 3(1), 25-42.

177. Stewart - Tull, D. (1970) "The setting and marking of microbiology examination", *Journal of Biological Education*, **4**(1), 221-235.

178. Sutcliffe, R., Leonard, E., Tierney, A., Howe, C., Reid, I., Goodwin, S. and Mackenzie, D. (2000) "Introduction of a range of computer - based objective tests in the examination of Genetics in first year biology", *Unpublished paper*, Glasgow, University of Glasgow.

179. Sutton, C. (1980) "The learner's prior knowledge: a critical review of techniques for probing its organization", *European Journal of Science Education*, **2**(2), 107-120.

180. Swain J. R. L. (1989) "The development of a framework for the assessment of process skills in a graded assessments in science project", *International Journal of Science Education*, **11**(3), 251-259.

181. Swain, J. R. L. (1991) "The nature and assessment of scientific exploration in the classroom", *School Science Review*, **72**(260), 65-77.

182. Tamir, P. (1990) "Justifying the selection of answers in multiple - choice items", *International Journal of Science Education*, **12**(5), 563-573.

183. Tan, K-C D. and Treagust, D. (1999) "Evaluating students' understanding of chemical bonding", *School Science Review*, **81**(294), 75-83.

184. Thyne, J. (1974) *Principles of Examining*, London, University of London Press LTD.

185. Tinajero, C. and P'aramo, M. (1997) "Field dependence - independence and academic achievement: a re-examination of their relationship", *British Journal of Educational Psychology*, **67**(2), 199-212.

186. Tinajero, C. and P'aramo, M. (1998) "Educational implications of field dependence - independence: in answer to Bagley and Mallick", *British Journal of Educational Psychology*, **68**, 589-593.

187. Tinajero, C., P'aramo, M., Cadaveira, F. and Rodriguez - Holguin, S. (1993) "Field dependence - independence and brain organization: The confluence of two different ways of describing general forms of cognitive functioning? A theoretical review", *Perceptual and Motor Skills*, 77, 787-802.

188. Torrance, J. (1996) *Standard Grade Biology*, London, Hodder and Stoughton.

189. Treagust, D. F. (1988) "Development and use of diagnostic tests to evaluate students' misconceptions in science", *International Journal of Science Education*, **10**(2), 159-169.

190. TRIADS Project Web Site (2000), The Web Site Address for TRIADS is: http://www.derby.ac.uk/assess/newdemo/mainmenu.html

191. Tulving, E, Schacter, D. and Stark, H. (1982) "Priming effects in word - fragment completion are independent of recognition memory", *Journal of Experimental Psychology: Learning, Memory and Cognition*, **8**(4), 336-342.

192. University of London (1976) "An introduction and guide to the use of multiple choice questions in university of examination", London, A document published by University of London.

193. Verma, M., Chatwal, J. and Singh, T.1997) "Reliability in essay type' questions - effect of structuring", *Assessment in Education*, 4(2), 265-270.

194. Waern, Y. (1972) "Structure in similarity matrices", *Scandinavian Journal of Psychology*, **13**, 5-16.

195. Wang, J. and Calhoun, G. (1997) "A useful function for assessing the effect of guessing on true - false and multiple choice tests", *Educational and Psychological Measurement*, **57**(1), 178-185.

196. Webster, A. and Walker, M (1981) "Divergent thinking in arts and science students: the effect of item content", *British Journal of Psychology*, **72**, 331-338.

197. West, D. (1993) "Introduction", in Edwards, D., Scanlon, E. and West, D. (eds.). *Teaching, Learning and Assessment in Science Education*, London, Paul Champman Publishing Ltd., 223-224.

198. White, R. and Gunstone, R. (1992) *Probing Understanding*, London, The Falmer Press.

199. Whitfield, R. C. (1970) "Criteria of quality for multiple - choice questions", *School Science Review*, **52**(178), 23-27.

200. Wills, D. (1993) "Learning and assessment: exposing the inconsistencies of theory and practice", *Oxford Review of Education*, **19**(3), 383-401.

201. Wiseman, S., and Pidgeon, D. (1970) *Exploring Education, Curriculum Evaluation*, London, National Foundation for Educational Research in England and Wales

202. Wisker, G. (1997) "Assessing for learning in English Studies: some innovative practices", *Teaching in Higher Education*, **2**(2), 123-139.

203. Witkin, H. and Goodenough, D. (1981) "Cognitive styles: essence and origin, field dependence and field independence", *Psychological Issues, Monograph 51*, New York, International Universities Press.

204. Witkin, H., Moore, C., Goodenough D. and Cox, P. (1977) "Field - dependent and field - independent cognitive styles and their educational implications", *Review of Educational Research*, 47(1), 1-64.

205. Zar, J. (1996) *Biostatistcal Analysis* (3ed. ed.) New Jersey, Prentice Hall Inc.

206. Zeidner, M. (1987) "Essay versus multiple - choice type classroom exams: the student's perspective", *Journal of Educational Research*, **80**(6), 352-358.

207. Ziane, M. (1990) "The application of information processing theory to the learning of physics", *Ph.D. Thesis*, Glasgow, University of Glasgow.

APPENDICES

APPENDIX (1): CHANGING THE POSITION OF THE MOST PLAUSIBLE DISTRACTOR (STRONG DISTRACTOR)

Appendix (1.1): Evidence for Evolution Test

Evidence For Evolution Test

1. All of the following statements are part of the Darwin-Wallace theory of natural selection <u>except</u>..................................

a. there exists in nature a constant struggle for survival.

b. characteristics which are acquired during the life of an individual are passed on to offspring.

c. organisms tend to increase in numbers at a rate more rapid than the environment can support.

d. heritable variations occur in natural populations.

e. on average the best adapted individuals leave more offspring.

2. During a study session about evolution one of your fellow student remarks, "The giraffe stretched its neck while reaching for higher leaves-----its offspring inherited longer necks as a result". Which of the following statements could you use to correct your fellow student's misconception?

a. overproduction of offspring leads to a struggle for survival.

b. disuse of an organ may lead to its eventual disappearance.

c. spontaneous mutations can result in the appearance of new traits.

d. only favorable adaptations have survival value.

e. characteristics acquired during an organism's life are not passed on through genes.

3. Darwin was able to formulate his theory of evolution based on several facts. Which of the following facts was unavailable to Darwin in the mid - nineteenth century?

a. most populations are stable in size.

b. all populations have the potential to increase.

c. individual organisms in a population are not alike.

d. natural resources are limited.

e. characteristics are inherited as genes on chromosomes.

4. What was the prevailing notion before Lyell and Darwin? The earth is......................

a. very old and populations gradually changed.

b. 6,000 years old and populations gradually change.

c. 6,000 years old and populations changed radically after periodic catastrophes.

d. 6,000 years old and populations are unchanged.

e. very old and populations were unchanging.

5. Which of the following is an acceptable definition of evolution?

a. a change in the species composition of a community.

b. a change in the environmental conditions.

c. a change in the genotypic makeup of an individual.

d. a change in the phenotypic makeup of a population.

e. a change in the genetic makeup of a population.

6. Of the following anatomical structures, which is homologous to the wing of a bat?

a. the tail fin of a fish.

b. the arm of the human.

c. the wing of a butterfly.

d. the dorsal fin of a shark.

e. the tail of a kangaroo.

7. Anatomic structures that show similar function but dissimilar embryonic and evolutionary background are said to be....................................

a. primitive.

b. polyphyletic.

c. homologous.

d. analogous.

e. monophyletic.

8. What are the ideas of Hutton and Lyell that Darwin incorporated into his theory?

a. adaptation of species to the environment.

b. the inheritance of acquired characteristics.

c. a hierarchical classification of organisms.

d. the age of Earth and gradual geological processes.

e. extinctions evident in the fossil record.

9. The gill slits of both reptile and bird embryos are..

a. primitive structures.

b. respiratory structures.

c. analogous structures.

d. vestigial structures.

e. homologous structures.

10. Natural selection is based on all of the following <u>except</u>......................................

a. there is differential reproductive success within populations.

b. the fittest individuals leave the most offspring.

c. populations tend to produce more individuals than the environment can support.

d. the need for individuals to adapt to their environment.

e. variation exists within populations.

11. Which of the following represents an idea Darwin took from the writings of Thomas Malthus?

a. the Earth is more than 10,000 years old.

b. all species are fixed in the form in which they are created.

c. the environment is responsible for natural selections.

d. the Earth changed over the years through a series of catastrophic upheavals.

e. populations tend to increase at a rate greater than their food supply.

12. All of the following disciplines have substantially contributed to the body of evidence of evolution <u>except</u>..................

a. biogeography.

b. mycology.

c. molecular biology.

d. taxonomy.

e. paleontology.

13. What was the most important missing evidence in Darwin's theory of natural selection?

a. that variation is common in populations.

b. that competition exists in populations.

c. that organisms became extinct.

d. the source of genetic variation.

e. that populations overproduce offspring.

14. Which of the following was <u>not</u> part of Darwin's explanation of natural selection?

a. variations exist within each species.

b. usually the most fit of each generation survive to reproduce.

c. members of a species compete with each other for food and space.

d. new variations continually arise by mutation.

e. organisms commonly produce more offspring than can possibly survive.

15. The statement "Improving the intelligence of an adult through education will results in that adult's descendants being born with a greater native intelligence", is an example of.........

a. Darwinism.

b. neo-Darwinism.

c. natural theology.

d. Lamarckism.

e. scala naturae.

16. Darwinism and Lamarckism both suggest that...

a. the Earth is 6,000 years old.

b. the environment creates favorable characteristics on demand.

c. species are fixed.

d. the interaction of organisms with their environment is important in the evolutionary process.

e. the main mechanism of evolution is the inheritance of acquired characteristics.

17. What did Charles Darwin publish in 1859?

a. Vestiges of Creation.

b. Philosophie Zoologique.

c. On the Nature of Things.

d. The Growth of Biological Thought.

e. On the Origin of Species by Means on Natural Selection.

18. The taxonomic system developed by Linnaeus is best described as a...................

a. map that distinguishes kinship among animals.

b. decimal plan for sorting all living organisms.

c. binary scheme of groupings.

d. hierarchy of increasingly general categories.

e. branching diagram of interrelationships.

19. Idealism, or essentialism, is an idea most associated with...

a. Lyell.

b. Plato.

c. Cuvier.

d. Darwin.

e. Lamarck.

20. Charles Lyell was an advocate of...

a. the modern synthesis.

b. uniformitarianism.

c. inheritance of acquired characteristics.

d. industrial melanism.

e. use and disuse.

21. Which of the following has provided undeniable evidence that the Earth has had a succession of flora and fauna?

a. the fossil record.

b. creationism.

c. natural selection.

d. population genetics.

e. catastrophism.

22. Which of the following would best describe Darwin?

a. catastrophist.

b. essentialist.

c. uniformitarianist.

d. gradualist.

e. creationist.

23. All of the following influenced Darwin as he synthesised the concept of natural selection <u>except</u>..................................

a. Lyell's "Principles of Geology".

b. the finches of the Galapagos.

c. the results of artificial selection.

d. Mendel's laws of inheritance.

e. Malthu's "Essays of Populations".

24. Who was the naturalist who synthesised a concept of natural selection independently of Darwin?

a. Mendal.

b. Malthus.

c. Lyell.

d. Wallace.

e. Henslow.

25. Which of the following elements of Darwinism is associated with Malthus?

a. favorable variations accumulate in a population after many generations of being perpetuated by natural selection.

b. artificial selection improves plant and animal breeds.

c. differential reproductive success is the cornerstone of natural selection.

d. species become better adapted to their local environments through natural selection.

e. the potential for population growth exceeds what the environment can support.

26. Which of the following concepts acknowledges the essential role of natural selection, stresses gradualism, and acknowledges the importance of the population as the unit of evolution?

a. ontogeny recapitulates phylogeny.

b. the modern synthesis.

c. ecological succession.

d. vitalism.

e. essentialism.

27. Current arguments by evolutionists about evolution are mainly concerned with the............

a. effects of ontogeny versus phylogeny.

b. mechanism of evolutionary change.

c. significance of natural versus artificial selection.

d. importance of homologous structures.

e. existence of vestigial organs.

28. Natural selection is based on all of the following except...

a. variation exists within populations.

b. populations tend to produce more individuals than the environment can support.

c. there is differential reproductive success with populations.

d. the environment tends to create favorable characteristics within populations.

e. the fittest individuals leave the most offspring.

29. The effect of natural selection in the case of the English peppered moth, ***Biston betularia***, illustrates that the advantage of inherited traits depends on the........................

a. intensity of melanin.

b. environment.

c. presence of homologies among moths.

d. principle of common descent.

e. presence of cytochrome c.

30. All of the following are facts or inferences of natural selection except...................................

a. unequal reproductive success leads to adaptations.

b. production of offspring is matched to the abundance of essential resources.

c. individuals whose inherited characteristics best fit them to the environment will leave more offspring.

d. since only a fraction of offspring survive, there is a struggle for limited resources.

e. there is heritable variation among individuals.

Appendix (1.2): Transport - Gas Exchange Test

Transport - Gas Exchange Test

1. Which of the following features do all gas exchange systems have in common?

a. they are enclosed within ribs.

b. they are found only in animals.

c. they are maintained at a constant temperature.

d. they are exposed to air.

e. they exchange surfaces are moist.

2. In which animal does blood flow through vessels from the respiratory organ back to the heart before circulating through the rest of the body?

a. frog.

b. fish.

c. annelid.

d. mollusc.

e. insect.

3. Which one of these statements about lungs is <u>false</u>?

a. the gases move across the exchange membranes by diffusion.

b. the total exchange surface area is large.

c. gas exchange takes place across moist membranes.

d. the concentration of CO_2 is higher in the air than in the alveolar capillaries.

e. the walls of the alveoli are only one cell thick.

4. At sea level, atmospheric pressure is 760 mm Hg. Oxygen gas constitute approximately 21% of the total gases in the atmosphere. What is the partial pressure of oxygen?

a. 21.0 mm Hg.

b. 76.0 mm Hg.

c. 160.0 mm Hg.

d. 508.0 mm Hg.

5. Which of one of the following animals would have the highest heart rate?

a. mouse.

b. elephant.

c. human.

d. cat.

e. horse.

6. Most of the carbon dioxide carried by the blood is carried as.......................................

a. bicarbonate attached to haemoglobin.

b. carbonic acid in the erythrocytes.

c. CO_2 attached to haemoglobin.

d. CO_2 dissolved in the plasma.

e. bicarbonate ion in the plasma.

7. What is the reason why fluid is forced out of systemic capillaries at the arteriolar end?

a. the somatic pressure of the interstitial fluid is greater than the hydrostatic pressure of the blood.

b. the hydrostatic pressure of the blood is greater than the osmotic pressure of the interstitial fluid.

c. the osmotic pressure of the blood is greater than the hydrostatic pressure of the interstitial fluid.

d. the osmotic pressure of the interstitial fluid is greater than that of the blood.

e. the hydrostatic pressure of the blood is less than that of the interstitial fluid.

8. If a molecule of CO_2 released into the blood in your left toe travels out of your nose, it must pass through all of the following structures <u>except</u> the....................................

a. right atrium.

b. pulmonary vein.

c. alveolus.

d. trachea.

e. right ventricle.

9. Which of the following is false concerning the haemoglobin molecule?

a. it is found in humans only.

b. it contains amino acids.

c. it is composed of four polypeptide chains.

d. it can bind four O_2 molecules.

e. it contains iron.

10. Which of the following conditions may predispose a person to a heart attack or stroke?

a. only thrombosis and high blood pressure (hypertension)

b. only atherosclerosis.

c. only thrombosis.

d. only high blood pressure (hypertension).

e. thrombosis, high blood pressure (hypertension), and atherosclerosis.

11. Which of the following sequences is not correct?

a. left ventricle----> aorta.

b. right ventricle---> pulmonary vein.

c. right ventricle---> pulmonary artery.

d. vena cava---> right atrium.

e. pulmonary vein---> left atrium.

12. Through how many capillary beds must human red blood cell travel if it takes the shortest possible route from the right ventricle to right atrium?

a. 1

b. 2

c. 3

d. 4

e. 5

13. What does it means if the blood pressure of a human is 110/80?

a. the diastolic pressure is 80 mm Hg.

b. the systolic pressure is 80 mm Hg.

c. the pulse rate is 80 beats per minute.

d. the pulse rate is 110 beats per minute.

14. Which of the following are the only vertebrates in which blood flows directly from the respiratory organs to the tissues without first returning to the heart?

a. amphibians.

b. fishes.

c. birds.

d. mammals.

e. reptiles.

15. Which of the following is <u>correct</u> for a blood pressure of 130/80?

a. the blood pressure during heart contraction is 80 mm Hg.

b. the diastolic pressure is 130 mm Hg and the systolic pressure is 80 mm Hg.

c. the systolic pressure is 130 mm Hg and the diastolic pressure is 80 mm Hg.

16. Which protein is the main constituent of the meshwork that forms the fabric of a blood clot?

a. fibrinogen.

b. fibrin.

c. thrombin.

d. prothrombin.

e. collagen.

17. If, during protein starvation, the osmotic pressure on the venous side of capillary beds drops below the hydrostatic pressure, then................................

a. haemoglobin will not release oxygen.

b. fluids will tend to accumulate in tissues.

c. plasma proteins will escape through the endothelium of the capillaries.

d. most carbon dioxide will not be carried away from tissues bound to haemoglobin.

e. the pH of the interstitial fluids will increase.

18. Which of the following is/are human plasma proteins?

a. immunoglobin.

b. haemoglobin.

c. fibrin.

19. Which of the following occurs with the exhalation of air from human lungs?

a. the volume of the theoretic cavity decreases.

b. the residual volume of the lungs decreases.

c. the diaphragm contracts.

d. the epiglottis closes.

e. the rib cage expands.

The following questions (20-21) refer to the data shown below in which the blood entering a vertebrate's capillary bed was measured for the pressures (in mm Hg) exerted by venous.

	Arterial End of Capillary Bed	**Venous End of Capillary Bed**
Hydrostatic Pressure	38 mm	14 mm
Osmotic Pressure	26 mm	26 mm
PO_2	100 mm	42 mm
PCO_2	40 mm	46 mm

20. For this capillary, which one of the following statements is correct?

a. the pH is lower on the arterial side than on the venous side.

b. oxygen is taken up by the erythrocytes within the capillaries.

c. the osmotic pressure remains constant due to carbon dioxide compensation.

d. the hydrostatic pressure declines from arterial to venous sides because oxygen is lost.

e. fluids will leave the capillaries on the arterial side of the bed and re-enter on the venous side.

21. The site of this capillary bed could be all of the following <u>except</u>........................

a. pancreas.

b. muscle tissue.

c. medulla.

d. alveoli.

e. kidneys.

Appendix (1.3): Kellett's Method of Calculating the Level of Significance

Kellett's Method of Calculating the Level of Significance

This is a stringent test that does not assume a normal distribution. The following are the steps showing how to calculate the level of significance.

1. Designate your two samples into smaller (N1) and larger (N2).
2. Calculate N1/N2 which gives you a value in the **X** - axis.
3. From the graph, which is shown in the next page, read off Δ (the value in the Y-axis) at both 5% and 1% levels. Then:
   - If ratio of X (N1/N2) ~ 0.1 or 0.9, you need to do the following:

     $$0.6 \times (\Delta/\sqrt{N1})$$

   - If ratio of X (N1/N2) ~ 0.2 or 0.8, you need to do the following:

     $$0.8 \times (\Delta/\sqrt{N1})$$

   - If ratio of X (N1/N2) ~ 0.3 or 0.7, you need to do the following:

     $$0.9 \times (\Delta/\sqrt{N1})$$

   - If ratio of X (N1/N2) ~ 0.4 or 0.6, you need to do the following:

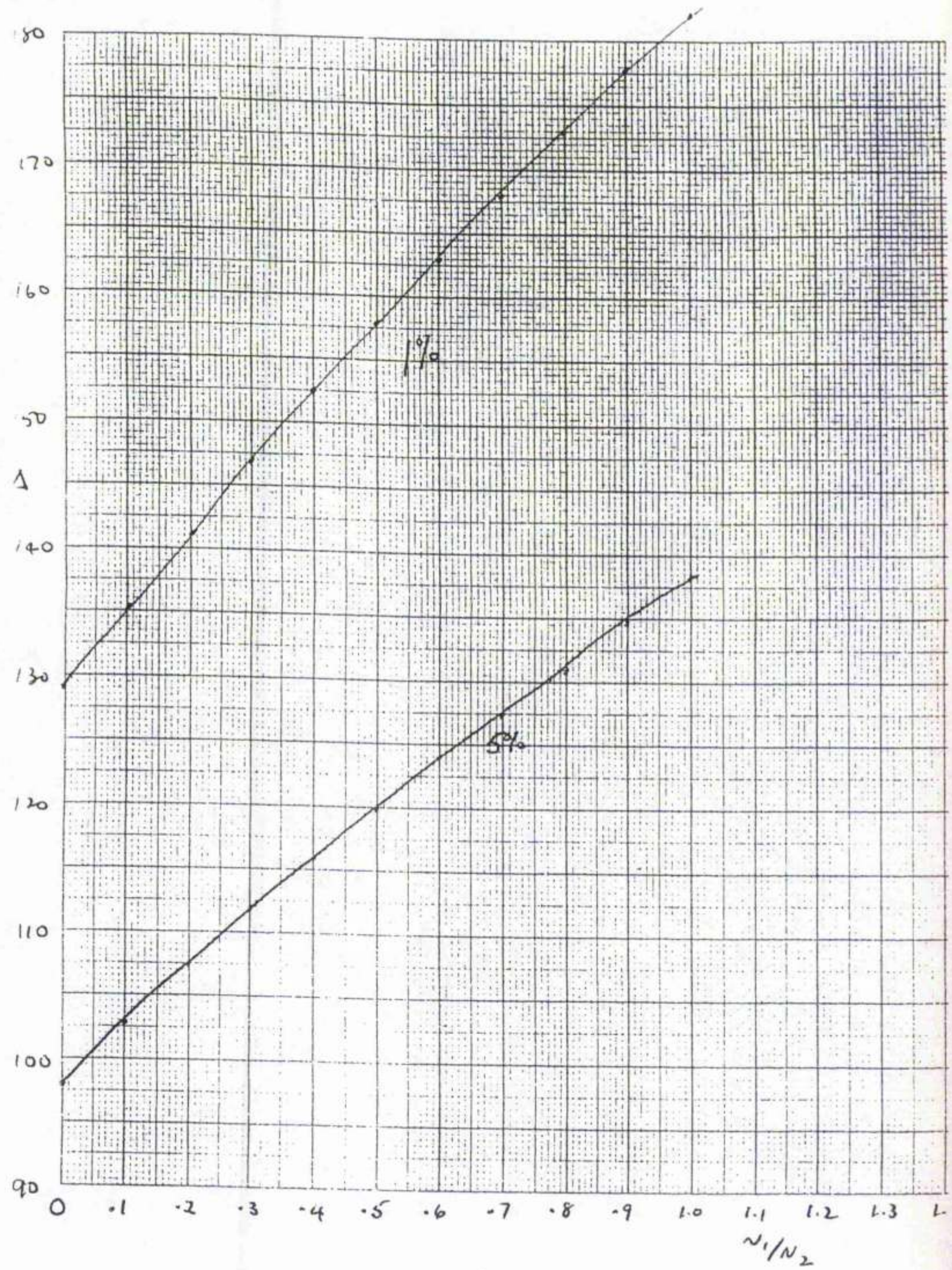
     $$0.98 \times (\Delta/\sqrt{N1})$$

   - If ratio of X (N1/N2) ~ 0.05 or 0.98, you need to do the following:

     $$0.43 \times (\Delta/\sqrt{N1})$$

4. Calculate the actual difference between the two facility values by subtracting the smaller value from the higher (F.V (higher) – F.V (smaller)).
5. In order that the difference in the facility value is significant, the actual difference (calculated in step **4**) must be larger than the value calculated in step **3**.

The Graph of Kellett's Method



APPENDIX (2): CHANGING THE POSITION OF THE KEY ANSWER

Appendix (2.1): Introduction to Animal Physiology Test

Introduction to Animal Physiology Test

1. What is stratified cuboidal epithelium composed of?

a. several layers of boxlike cells.

b. a hierarchical arrangement of flat cells.

c. a tight layer of square cells attached to a basement membrane.

d. an irregularly arranged layer of pillarlike cells.

e. a layer of ciliated, mucus-secreting cells lining a body cavity.

2. Interstitial fluid...

a. forms the extracellular matrix surrounding the cells of connective tissue.

b. composed of blood.

c. provides for the exchange of materials between blood and cells.

d. is the liquid portion of blood.

3. In a typical multicellular animal, the circulatory system interacts with specialized surfaces in order to exchange materials with the exterior environment. Which of the following is not an example of such an exchange surface?

a. lung.

b. muscle.

c. intestine.

d. kidney.

4. Stratified columnar is a description that might apply to what type of animal tissue?

a. connective.

b. striated muscle.

c. nerve.

d. epithelial.

e. bone.

5. The epithelium best adapted for a body surface subject to abrasion is...........................

a. simple squamous.

b. simple cuboidal.

c. simple columnar.

d. stratified columnar.

e. stratified squamous.

6. Muscles are joined to bones by...

a. ligaments.

b. tendons.

c. loose connective tissue.

d. Haversian systems.

7. Fibroblasts secrete...

a. fats.

b. chondrin.

c. interstitial fluids.

d. calcium phosphate for bone.

e. proteins for connective fibers.

8. Which type of muscle is responsible for peristalsis along the digestive tract?

a. cardiac.

b. smooth.

c. voluntary.

d. striated.

e. skeletal.

9. In mammals, the diaphragm separates the abdominal cavity from the............................

a. coelom.

b. pharynx.

c. thoracic cavity.

d. gastrovascular cavity.

e. oral cavity.

10. Which of the following is a direct problem that had to be solved for animals to increase in size?

I. Decreasing surface- to - volume ratio.

II. Reproducing in aqueous environments.

III. Increasing tendency for larger bodies to be more variable.

a. I only.

b. II only.

c. III only.

d. I and III only.

e. I, II, III.

11. Cardiac muscle is which of the following?

a. striated and involuntary.

b. smooth and voluntary.

c. striated and voluntary.

d. smooth and involuntary.

12. Which of the following tissues lines kidney ducts?

a. connective.

b. smooth muscle.

c. nervous.

d. epithelial.

e. adipose.

13. Which of the following descriptions applies to skeletal muscle?

a. smooth and involuntary.

b. striated and voluntary.

c. smooth and voluntary.

d. striated and involuntary.

14. Cartilage is described as which of the following types of tissues?

a. connective.

b. reproductive.

c. nervous.

d. epithelial.

e. adipose.

15. Bones are joined together at joints by...

a. negative feedback.

b. Haversian systems.

c. loose connective tissues.

d. tendons.

e. ligaments.

16. Which of the following fibers has the greatest tensile strength?

a. elastin fibers.

b. fibrin fibers.

c. collagenous fibers.

d. reticular fibers.

e. spindle fibers.

17. Which of the following is an example of positive feedback?

a. an increase in blood sugar concentration increases the amount of the hormone that stores sugar as glycogen.

b. a decrease in blood sugar concentration increases the amount of hormone that converts glycogen to glucose.

c. an infant's suckling at the mother's breast increases the amount of the hormone that includes the release of milk from the mammary glands.

d. an increase in calcium concentration increases the amount of the hormone that stores calcium in bone.

e. a decrease in calcium concentration increases the amount of hormone that releases calcium from the bone.

18. Why must multicellular organisms keep their cells awash in an 'internal pond'?

a. negative feedback will only operate in interstitial fluids.

b. all cells need an aqueous medium for the exchange of food, gases and wastes.

c. the cells of multicellular organisms tend to lose water because of osmosis.

d. the cells of multicellular organisms tend to accumulate wastes, a consequence of diffusion.

e. this phenomenon only occurs in aquatic organisms because terrestrial organisms have adapted to life in a dry situation.

Appendix (2.2): Origin of Species Test

Origin of Species Test

1. The only taxonomic category that actually exists as a discrete unit in nature is the..............................

a. class.

b. species.

c. family.

d. genus.

e. phylum.

2. To a punctuationalist, the "sudden" appearance of a new species in the fossil record means that..

a. the species is now extinct.

b. speciation occurred instantaneously.

c. the Earth is only 6000 years old.

d. speciation occurred in one generation.

e. speciation occurred over many thousand years.

3. The biologist who proposed the biological species concept is..........................

a. Mayr.

b. Sheldon.

c. Wright.

d. Gould.

e. Eldridge.

4. Which of the following statements is consistent with the punctuated equilibrium interpretation of speciation?

a. evolution proceeds at a slow, steady pace.

b. rapid speciation is caused by population explosions.

c. large populations evolve more quickly than small ones.

d. long periods of minor change are interrupted by short bursts of significant change.

e. there is an equilibrium between living and extinct species.

5. A rapid method of speciation that has been important in the history of flowering plants is...

a. parapatric speciation.

b. genetic drift.

c. a mutation in the gene controlling the timing of flowering.

d. polyploidy.

e. behavioral isolation.

6. If two species are able to interbreed but produce sterile hybrids, their species integrity is maintained by......................

a. hybrid inviability.

b. hybrid breakdown.

c. a postzygotic barrier.

d. gametic isolation.

e. a prezygotic barrier.

7. A botanist discovers a large population of annual plants. The plants all look basically the same, but seem to be of two different size classes. The larger and smaller plants inhabit the same areas and are visited by the same pollinating insects. What is the most likely reason for the size differences?

a. the larger plants are dominant and the smaller plants are recessive.

b. the larger plants happen to be in areas with more nutrients.

c. the larger plants are polyploids derived from the smaller plants.

d. the smaller plants are haploids which developed from unfertilized eggs.

e. the larger plants germinated in the winter and the smaller plants germinated in the summer.

8. A botanist discovers a large population of annual plants. The plants all look basically the same, but seem to be of two different size classes. The larger and smaller plants inhabit the same areas and are visited by the same pollinating insects. What would be the <u>least</u> productive research to discover the relationship between the plants?

a. growing seed from the two sizes of plants under identical conditions.

b. growing seeds of one size of plant with various nutrient concentrations.

c. careful measurement of anatomical features.

d. chromosomes counts.

e. electrophorectic studies to see if they have the same enzymes.

9. A botanist discovers a large population of annual plants. The plants all look basically the same, but seem to be of two different size classes. The larger and smaller plants inhabit the same areas and are visited by the same pollinating insects. Under what conditions might the two sizes of plants be considered different species?

a. microscopy shows the bigger plants have bigger cells.

b. pollinating insects prefer the larger plant to the smaller.

c. pollen from plants fails to germinate on the stigmas of plants of the other size class.

d. tests show genetic differences between the plants.

e. a cross between the two sizes of plants yields hybrids of intermediate size.

10. Which of the following is not considered to be a reproductive isolating mechanism?

a. gametic incompatibility.

b. ecological isolation.

c. sterile offspring.

d. timing of courtship display.

e. feeding behavior.

11. Which of the following reproductive isolating mechanisms is postzygotic?

a. hybrid sterility.

b. gamete incompatibility.

c. habitat isolation.

d. temporal isolation.

e. behavioral isolation.

12. Some species of *Anopheles* mosquito live in brackish water, some in running fresh water, and others in stagnant water. What type of reproductive barrier is most obviously separating these different species?

a. temporal isolation.

b. ecological isolation.

c. gametic isolation.

d. postzygotic isolation.

e. behavioral isolation.

13. According to advocates of the punctuated equilibrium theory,.........................

a. natural selection is unimportant as a mechanism of evolution.

b. a new species accumulates most of its unique features as it comes into existence.

c. most evolution results from a disruption of Hardy-Weinberg equilibrium.

d. transitional fossils are intermediate between newer species and their parent species.

e. given enough time, most existing species will branch gradually into new species.

14. The biological species concept is inadequate for grouping............................

a. parasites.

b. endemic populations.

c. sympatric populations.

d. plants.

e. asexual organisms.

15. A characteristic of allopatric speciation is..

a. the appearance of new species in the midst of old ones.

b. large populations.

c. geographic isolation.

d. artificial selection.

e. asexually reproducing populations.

16. The process of a new species arising within the range of the parent populations is termed..................................

a. parapatric speciation.

b. allopatric speciation.

c. sympatric speciation.

d. adaptive radiation.

e. semispeciation.

17. The biologists who proposed the theory of punctuated equilibrium are................................

a. Sheldon and Templeton.

b. Darwin and Wallace.

c. Mayr and Wright.

d. Eldredge and Gould.

e. Watson and Crick.

18. The origin of a new plant species by hybridization coupled with nondisjunction is an example of...

a. autopolyploidy.

b. sympatric speciation.

c. introgression.

d. a peak shift.

e. allopatric speciation.

19. All of the following statements about 'splinter' populations, or peripheral isolates, are correct <u>except</u>...

a. the selective factors operating on peripheral isolates may be quite different from those operating on the parent populations.

b. life on the frontier is usually harsh for the peripheral isolates and most become extinct.

c. the gene pool may represent the extremes of genotypic and phenotypic clines.

d. many peripheral isolates have an increased likelihood of experiencing a founder effect.

e. they undergo speciation readily because they are large populations with immense gene pools.

20. Which of the following would be a position held by a punctuationalist?

a. given enough time, most existing species will gradually evolve into new species.

b. natural selection is unimportant as a mechanism of evolution.

c. one should expect to find many transitional fossils left by organisms in the process of forming new species.

d. most speciation is anagenic.

e. a new species forms most of its unique features as it comes into existence, and then changes little for the duration of its existence.

21. The Hawaiian Islands are the greatest showcase of evolution because of intense...................

a. cross -specific mating and reinforcement.

b. ecological isolation and parapatric speciation.

c. adaptive radiation and allopatric speciation.

d. allopolyploidy and sympatric speciation.

e. hybrid vigor and allopatric speciation.

APPENDIX (3): EXAMINE THE 'OPTICAL' EXPLANATION BEHIND THE DIFFERNCES IN THE FACILITY VALUES OF THE MULTIPLE CHOICE QUESTIONS

FORM (A) TEST

REVISION EXERCISES

CHEMISTRY STANDARD GRADE - FORM (A)

NAME:..

INSTRUCTIONS

1. This test is for practice only. The results will not affect your academic work or exam in any way.
2. The questions are in the multiple choice format.
3. All questions can be answered by drawing a circle around the appropriate letter in each question.
4. Questions are printed on both sides of the paper.
5. Please attempt as many questions as you can.
6. Do not start the test until your teacher tells you.

Q1. Which one of the following elements has an electron arrangement of 2, 8, 2?

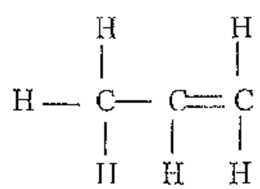
a. oxygen.

b. calcium.

c. magnesium.

d. fluorine.

Q2. Which one of the following hydrocarbons has the following structural formula?



a. propane.

b. pentane.

c. pentene.

d. propene.

Q3. Which one of the following changes is NOT an example of a chemical reaction?

a. breaking glass.

b. iron rusting.

c. methane burning.

d. neutralising acid.

Q4. Which one of the following elements conducts electricity at room temperature?

a. bromine.

b. helium.

c. mercury.

d. sulphur.

Q5. Which one of the following elements is an unreactive gas?

a. chlorine.

b. mercury.

c. sodium.

d. helium.

Q6. Which of the following pairs of elements have similar chemical properties?

a. bromine and chlorine.

b. helium and sodium.

c. mercury and sulphur.

d. sodium and sulphur.

Q7. Which one of the following compounds is produced when Nitric Acid is neutralised by an alkali?

a. calcium iodide.

b. potassium sulphate.

c. lithium nitrate.

d. sodium phosphate.

Q8. Which one of the following compounds is covalent?

a. calcium iodide.

b. phosphorus chloride.

c. potassium sulphate.

d. sodium phosphate.

Q9. Which one of the following compounds can produce a solution with a pH value of 11 when dissolved in water?

a. sodium hydroxide.

b. hydrochloric acid.

c. carbonic acid.

d. sodium chloride.

Q10. Which one of the following elements is a halogen?

a. chlorine.

b. calcium.

c. mercury.

d. magnesium.

FORM (B) TEST

REVISION EXERCISES

CHEMISTRY STANDARD GRADE - FORM (B)

NAME:..

INSTRUCTIONS

1. This test is for practice only. The results will not affect your academic work or exam in any way.
2. All questions are in the multiple choice format.
3. In order to answer each question, you need first to select the two options you think are definitely wrong by pulling them from the test and throw them away. Now you have two options remaining in the question. Draw a circle around the option you think is the correct one.
4. Please attempt as many questions as you can.
5. Do not start the test until your teacher tells you.

Q1. Which one of the following elements has an electron arrangement of 2, 8, 2?

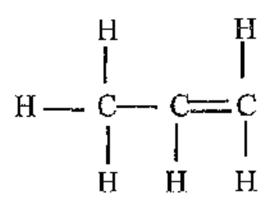
a. oxygen.

b. calcium.

c. magnesium.

d. fluorine.

Q2. Which one of the following hydrocarbons has the following structural formula?



a. propane.

b. pentane.

c. pentene.

d. propene.

Q3. Which one of the following changes is NOT an example of a chemical reaction?

a. breaking glass.

b. iron rusting.

c. methane burning.

d. neutralising acid.

Q4. Which one of the following elements conducts electricity at room temperature?

a. bromine.

b. helium.

c. mercury.

d. sulphur.

Q5. Which one of the following elements is an unreactive gas?

a. chlorine.

b. mercury.

c. sodium.

d. helium.

Q6. Which of the following pairs of elements have similar chemical properties?

a. bromine and chlorine.

b. helium and sodium.

c. mercury and sulphur.

d. sodium and sulphur.

Q7. Which one of the following compounds is produced when Nitric Acid is neutralised by an alkali?

a. calcium iodide.

b. potassium sulphate.

c. lithium nitrate.

d. sodium phosphate.

Q8. Which one of the following compounds is covalent?

a. calcium iodide.

b. phosphorous chloride.

c. potassium sulphate.

d. sodium phosphate.

Q9. Which one of the following compounds can produce a solution with a pH value of 11 when dissolved in water?

a. sodium hydroxide.

b. hydrochloric acid.

c. carbonic acid.

d. sodium chloride.

Q10. Which one of the following elements is a halogen?

a. chlorine.

b. calcium.

c. mercury.

d. magnesium.

APPENDIX (4): CHANGING THE POSITIONS OF THE OPTIONS WITHIN THE GRID

Appendix (4.1): The School Level Test

REVISION EXERCISE

(Chemistry Standard Grade)

INSTRUCTIONS

1. This test is for practice only. The results will not affect your academic work or exam in any way.

2. All questions can be answered by drawing a circle around the appropriate letter (or letters) in the answer grid provided for each question.

3. Please attempt as many questions as you can.

Q1. Jim was asked to find out whether fertilisers containing Potassium ion K^+ or fertilisers containing Ammonium ions NH_4^+ are better for growing lettuces. He made up two fertiliser solutions for his experiment.

(a) Select the TWO solutions which Jim could have used in a fair test?

KNO_3 concentration 1 mol/l **A**	K_2SO_4 concentration 1 mol/l **B**	K_2SO_4 concentration 2 mol/l **C**
NH_4Cl concentration 1 mol/l **D**	$(NH_4)_2SO_4$ concentration 1 mol/l **E**	NH_4NO_3 concentration 2 mol/l **F**

A	B	C
D	E	F

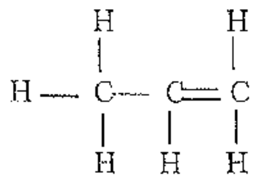
Q2. Hydrocarbon compounds are obtained from crude oil.

Butene **A**	Propane **B**	Hexene **C**
Pentane **D**	Pentene **E**	Propene **F**

(a) Select the TWO hydrocarbons which are Alkanes?

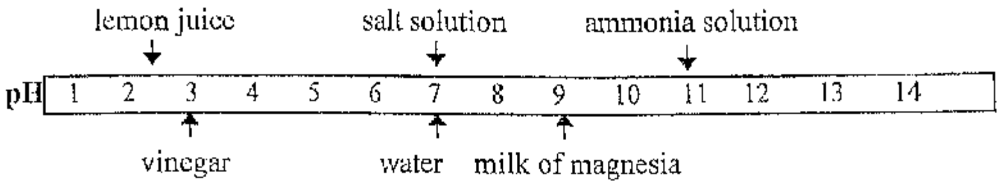
A	B	C
D	E	F

(b) Select the hydrocarbons which has the following structural formula?



A	B	C
D	E	F

Q3. The chart shows the pH of some common substances.



Ammonia Solution **A**	Lemon Juice **B**	Milk of Magnesia **C**
Salt Solution **D**	Vinegar **E**	Water **F**

(a) Select the TWO substances which are acids?

A	B	C
D	E	F

(b) Select the TWO substances which will show a DECREASES in pH when they are diluted with water?

A	B	C
D	E	F

Q4. Chemical reactions make new substances.

Ice melting A	Neutralising acid B	Iron rusting C
Breaking glass D	Methane burning E	Polymerising ethene F

(a) Select the TWO changes which are NOT examples of chemical reactions?

A	B	C
D	E	F

(b) Select the chemical reaction(s) which would produce water as one of the products?

A	B	C
D	E	F

Q5. Elements have different properties.

Bromine A	Chlorine B	Helium C
Mercury D	Sodium E	Sulphur F

(a) Select the TWO elements which conduct electricity at room temperature?

A	B	C
D	E	F

(b) Select unreactive gas?

A	B	C
D	E	F

(C) Select the TWO non-metal elements which have similar chemical properties?

A	B	C
D	E	F

Q6. There are many different chemical compounds.

Lithium Nitrate A	Calcium Iodide B	Magnesium Carbonate C
Phosphorus Chloride D	Potassium Sulphate E	Sodium Phosphate F

(a) Select the compound produced when Nitric Acid is neutralised by an Alkali?

A	B	C
D	E	F

(b) Select the covalent compound?

A	B	C
D	E	F

Q7. There are many different chemical compounds.

$CH_3-CH-CH_2-CH_3$ $\quad\ \ |$ $\quad CH_3$ A	$CH_3-CH=CH-CH_3$ B	$CH_3-CH-CH_3$ $\qquad\ |$ $\quad\ \ CH_3$ C
CH_2-CH_2 $\ \ |\qquad\ \ |$ CH_2-CH_2 D	$CH_3-CH_2-CH_3$ E	CH_4 F

(a) Select the hydrocarbon with a formula mass of 58 amu?

A	B	C
D	E	F

(b) Select the TWO hydrocarbons which are NOT in the same homologous series as C_8H_{18} ?

A	B	C
D	E	F

Q8. Ion-electron equations can be used to show the gain and loss of electrons in chemical reactions.

$Fe_{(s)} \rightarrow Fe^{2+}{}_{(aq)} + 2e$ **A**	$Fe^{2+}{}_{(aq)} + 2e \rightarrow Fe_{(s)}$ **B**	$Fe^{2+}{}_{(aq)} \rightarrow Fe^{3+}{}_{(aq)} + e$ **C**
$Fe^{3+}{}_{(aq)} + e \rightarrow Fe^{2+}{}_{(aq)}$ **D**	$Cu_{(s)} \rightarrow Cu^{2+}{}_{(aq)} + 2e$ **E**	$Cu^{2+}{}_{(aq)} + 2e \rightarrow Cu_{(s)}$ **F**

(a) Select the equation which shows iron (II) ions being oxidised ?

A	B	C
D	E	F

(b) Select the TWO equations which show the reactions which occur when an iron nail is placed in copper (II) sulphate solutions?

A	B	C
D	E	F

Appendix (4.2): Pupils' Ranking in Two Scoring Methods of the Grid Questions

Pupils' Ranking in Scottish Qualifications Authority Scoring Method (S.Q.A.S.M) and the Method Suggested by Egan

	S.Q.A.S.M		Egan Method			S.Q.A.S.M		Egan Method	
Student	Score	Rank	Score	Rank	Student	Score	Rank	Score	Rank
1	0	1	1.12	3	28	6	31	5.05	28
2	3	3.5	0.21	1	29	6	31	5.37	29
3	3	3.5	1.06	2	30	6	31	5.47	30
4	3	3.5	1.65	4	31	6	31	5.68	31
5	3	3.5	2.12	5	32	6	31	5.79	32.5
6	4	10	2.13	6	33	6	31	5.79	32.5
7	4	10	2.39	7	34	6	31	6.11	34
8	4	10	2.44	8	35	6	31	6.12	35
9	4	10	2.92	9	36	7	39	6.27	36
10	4	10	3.13	10	37	7	39	6.32	37.5
11	4	10	3.45	11.5	38	7	39	6.32	37.5
12	4	10	3.93	15	39	7	39	6.38	39
13	4	10	4.03	16	40	7	39	6.80	40
14	4	10	4.20	19	41	7	39	6.85	41
15	5	20.5	3.45	11.5	42	7	39	6.91	42
16	5	20.5	3.72	13	43	8	47.5	7.07	43.5
17	5	20.5	3.88	14	44	8	47.5	7.07	43.5
18	5	20.5	4.09	17	45	8	47.5	7.12	45.5
19	5	20.5	4.20	19	46	8	47.5	7.12	45.5
20	5	20.5	4.20	19	47	8	47.5	7.38	48.5
21	5	20.5	4.25	21	48	8	47.5	7.38	48.5
22	5	20.5	4.46	22.5	49	8	47.5	7.38	48.5
23	5	20.5	4.46	22.5	50	8	47.5	7.38	48.5
24	5	20.5	4.52	24	51	8	47.5	7.44	51
25	5	20.5	4.89	26	52	8	47.5	7.60	52.5
26	5	20.5	4.99	27	53	9	62	7.60	52.5
27	6	31	4.78	25	54	9	62	7.81	54

Pupils' Ranking in Scottish Qualifications Authority Scoring Method (S.Q.A.S.M) and the Method Suggested by Egan

	S.Q.A.S.M		Egan Method			S.Q.A.S.M		Egan Method	
Student	Score	Rank	Score	Rank	Student	Score	Rank	Score	Rank
55	9	62	8.18	55.5	82	10	79	9.93	81.5
56	9	62	8.18	55.5	83	10	79	9.99	83
57	9	62	8.55	57	84	10	79	10.20	85.5
58	9	62	8.61	58	85	10	79	10.25	87.5
59	9	62	8.71	60	86	10	79	10.25	87.5
60	9	62	8.71	60	87	11	95	10.15	84
61	9	62	8.71	60	88	11	95	10.20	85.5
62	9	62	8.87	62	89	11	95	10.41	89
63	9	62	8.93	63	90	11	95	10.47	90.5
64	9	62	8.98	64	91	11	95	10.47	90.5
65	9	62	9.14	66	92	11	95	10.52	92
66	9	62	9.14	66	93	11	95	9.14	93
67	9	62	9.14	66	94	11	95	10.68	93
68	9	62	9.19	68.5	95	11	95	10.78	94.5
69	9	62	9.46	72	96	11	95	10.78	94.5
70	9	62	9.46	72	97	11	95	11	96.5
71	9	62	9.62	76	98	11	95	11	96.5
72	10	79	9.19	68.5	99	11	95	11.05	99.5
73	10	79	9.46	72	100	11	95	11.05	99.5
74	10	79	9.46	72	101	11	95	11.05	99.5
75	10	79	9.46	72	102	11	95	11.05	99.5
76	10	79	9.51	75	103	11	95	11.26	104
77	10	79	9.67	77.5	104	12	112	11.26	104
78	10	79	9.67	77.5	105	12	112	11.26	104
79	10	79	9.72	79	106	12	112	11.53	107
80	10	79	9.78	80	107	12	112	11.53	107
81	10	79	9.93	81.5	108	12	112	11.53	107

Pupils' Ranking in Scottish Qualifications Authority Scoring Method (S.Q.A.S.M) and the Method Suggested by Egan

	S.Q.A.S.M		Egan Method			S.Q.A.S.M		Egan Method	
Student	**Score**	**Rank**	**Score**	**Rank**	**Student**	**Score**	**Rank**	**Score**	**Rank**
109	12	112	11.58	110	136	14	144	13.12	132
110	12	112	11.58	110	137	14	144	13.33	134.5
111	12	112	11.58	110	138	14	144	13.60	137
112	12	112	11.74	112	139	14	144	13.65	139.5
113	12	112	11.79	113	140	14	144	13.65	139.5
114	12	112	11.85	114	141	14	144	13.87	141
115	12	112	12.01	115	142	14	144	13.92	143
116	12	112	12.06	116.5	143	14	144	13.92	143
117	12	112	12.06	116.5	144	14	144	13.92	143
118	12	112	12.33	120	145	14	144	14.13	145
119	12	112	12.33	120	146	14	144	14.34	146
120	12	112	12.54	124	147	14	144	14.40	145.5
121	13	128	12.27	118	148	14	144	14.40	145.5
122	13	128	12.33	120	149	14	144	14.40	145.5
123	13	128	12.38	122.5	150	14	144	14.45	148
124	13	128	12.38	122.5	151	14	144	14.45	148
125	13	128	12.80	125	152	14	144	14.61	152
126	13	128	12.86	128.5	153	15	156.5	14.66	153
127	13	128	12.86	128.5	154	15	156.5	14.93	155
128	13	128	12.86	128.5	155	15	156.5	14.93	155
129	13	128	12.86	128.5	156	15	156.5	14.93	155
130	13	128	12.86	128.5	157	15	156.5	15.14	157
131	13	128	12.86	128.5	158	15	156.5	15.19	159
132	13	128	13.18	133	159	15	156.5	15.19	159
133	13	128	13.33	134.5	160	15	156.5	15.19	159
134	13	128	13.60	137	161	16	164	15.73	162.5
135	13	128	13.60	137	162	16	164	15.73	162.5

Pupils' Ranking in Scottish Qualifications Authority Scoring Method (S.Q.A.S.M) and the Method Suggested by Egan

	S.Q.A.S.M		Egan Method			S.Q.A.S.M		Egan Method	
Student	**Score**	**Rank**	**Score**	**Rank**	**Student**	**Score**	**Rank**	**Score**	**Rank**
162	16	164	15.73	162.5	168	17	170.5	17	170.5
163	16	164	15.73	162.5	169	17	170.5	17	170.5
164	16	164	16.47	166	170	17	170.5	17	170.5
165	16	164	16.47	166	171	17	170.5	17	170.5
166	16	164	16.47	166	172	17	170.5	17	170.5
167	17	170.5	17	170.5					

Appendix (4.3): The University Level Test

REVISION EXERCISE

(FIRST YEAR UNIVERSITY CHEMISTRY)

Name:---**Matriculation No.:**-------------------------

INSTRUCTIONS

1. This test is a revision exercise. The results will not affect your academic work or exam in any way but it will be very useful for the forthcoming test.
2. All sub-questions are based upon the grid provided in each question.
3. The answer will require a set of numbers corresponding to the box numbers in the grid. For each sub-question, you need to state the box(es) chosen (their numbers) from the grid.
4. Questions are printed on both sides of the paper.
5. Please attempt as many questions as you can.

Question One:

$^{17}_{35}Cl$ 1	$^{55}_{132}Cs$ 2	$^{53}_{126}I$ 3	$^{3}_{6}Li$ 4
$^{6}_{12}C$ 5	$^{9}_{18}F$ 6	$^{10}_{20}Ne$ 7	$^{13}_{26}Al$ 8

a. Which box contains the element with the **smallest** ionization energy?

____

b. Which box contains the element with the **largest** electron affinity?

____

Question Two:

$^{25}_{54}Mn$ 1	$^{17}_{35}Cl$ 2	$^{3}_{6}Li$ 3	$^{35}_{79}Br$ 4
$^{8}_{15}O$ 5	$^{5}_{10}B$ 6	$^{14}_{28}Si$ 7	$^{9}_{18}F$ 8

a. Which box(es) contain(s) metals?

____

b. Which box(es) contain(s) metalloids?

____

c.i Select the box(es) contain(s) halogens ?

____

ii. Order your selections in **c.i** above according to their **increase** in ionization energy?

____

d.i Select the box(es) contain(s) elements in the second row of the periodic table?

____

ii. Order your selections in **d.i** above according to their **decrease** in electronegativivty?

____

Question Three:

${}^{12}_{24}Mg$ 1	${}^{1}_{1}H$ 2	${}^{2}_{4}He$ 3	${}^{28}_{58}Ni$ 4
${}^{4}_{9}Be$ 5	${}^{20}_{40}Ca$ 6	${}^{21}_{44}Sc$ 7	${}^{16}_{32}S$ 8

a. Which box contains element which has only ns^2 electron configuration?

____

b. Which box contains element with $[Ar]4s^2$ electron configuration?

____

c. Which box contains element which is located in column two and row three of the periodic table?

____

d. Which box(es) contain(s) element(s) which is (are) located in the same column of the periodic table? ______________________________

e. Which box(es) contain(s) element(s) in which the d subshell contains an electron(s)?

____

Question Four:

$Ca^{2+} \rightarrow Ca$ 1	$Cr_2O_7^{2-}$ 2	$HClO_4$ 3	Cl_2 4
H_2O_2 5	ClO_3^- 6	CrO_4^{2-} 7	$Sn^{2+} \rightarrow Sn^{4+}$ 8

a. Which box(es) contain(s) chlorine (**Cl**) in the oxidation state of +5?

___ ___ ___ ___ ___ ___ ___ ___

b. Which box(es) contain(s) the chemical equation which shows oxidation?

___ ___ ___ ___ ___ ___ ___ ___

c. Which box(es) contain(s) chromium (**Cr**) in the oxidation state of +6?

___ ___ ___ ___ ___ ___ ___ ___

d. Which box(es) contain(s) oxygen (**O**) in the oxidation state of -2?

___ ___ ___ ___ ___ ___ ___ ___

Question Five:

MgF_2 1	${}^{4}_{9}Be$ 2	F_2 3	NH_4Cl 4
C_2H_4 5	Cl_2 6	${}^{26}_{55}Fe$ 7	C_2H_3Cl 8

a. Which box(es) contain(s) ionic compounds?

____

b. Which box(es) contain(s) species which has(ve) both ionic and covalent bonding?

____

c. Which box(es) contain(s) species which has(ve) double bonds?

____

d. Which box(es) contain(s) species which could form a covalent compound with carbon (C)?

____

e. Which box(es) contain(s) a nonpolar species?

____

Question Six:

NH_3 1	ClF_2^- 2	square planar 3	ClF_3 4
square pyramidal 5	BeF_2 6	BrF_3 7	I_3^- 8

a. Which box(es) contain(s) species which has(ve) a trigonal pyramidal molecular geometry ?

____

b. Which box(es) contain(s) species which has(ve) a linear molecular geometry?

____

c. Which box(es) contain(s) species which has(ve) a **T** shaped molecular geometry ?

____

d. Which box(es) contain(s) the type of molecular geometry of IF_5 and BrF_5 ?

____

APPENDIX (5): COMPUTER PROGRAM FOR DESIGNING AND SCORING GRID QUESTIONS

Appendix (5.1): The User Manual

COMPUTER PROGRAMME FOR SCGs)

(User Manual)

Introduction:

An attempt has been made to design a computer program to provide shells in a grid form as a tool for self - assessment and to speed up the marking process of SCGs. The program was designed by Allan Lucas who was an undergraduate student in the Department of Computing Science – University of Glasgow. This program was designed to work in any personal computer (**PC**) which has **Windows 95** and onwards. It was designed to score students' selection(s) according to the method suggested by Egan (Coefficient of Confusion) which can be calculated according to the following formula:

Cofficient of Confusion (Scores) =

$$\frac{\text{No. of relevant (✓) responses chosen}}{\text{No. of relevant (✓) responses available}} - \frac{\text{No. of irrelevant (✗) responses chosen}}{\text{No. of irrelevant (✗) responses available}}$$

Mackenize's third method of scoring (**Before and Next to**) was used to score the logical order of students' selection(s).

How to Install the Program in your PC:

The program can be installed in one central computer (i.e. Computer Server) if the institution has a computer network. There are two ways in which the user can install the program in his computer:

1. The program is given a name called 'MARKSYS'. Before installing MARKSYS, please note that the automatic installation (Install + Run icon) will only work if you have a HARD DRIVE called c or C and that you have a CR-ROM called d or D (i.e. CD-ROM Drive). If so then double click on the **jre12I-win32** icon. This will start the installation of the JAVA runtime environment, which is needed to run MARKSYS. After double clicking on the jre12l-win32 icon a window will appear. Follow the instruction in the window to install the JAVA runtime environment. Once this is done, double click on the Install + Run icon. This will install the MARKSYS program into your c or C drive (Hard Drive).

2. If the auto installation does not work (because of drive letters as mentioned above), then manually copy the system folder from the CD to your hard drive (double click on your CD - ROM icon. Right click on the system folder, select copy. Double click on your Hard Drive. Right click in your Hard Drive window and select paste). Double click on the **jre12I-win32** icon and install the Java runtime environment. The MARKSYS program can then be run by entering the system folder (the one you just copied and on your Hard Drive not the one on you CD - ROM). Then double click on the Run icon.

After the installation takes place, the user can run the program any time either by entering the MARKSYS folder (this if you installed MARKSYS program according to installation note 1) which will be on your c or C HARD DRIVE. Then double click on the Run icon. OR by entering the system folder (if you installed MARKSYS program according to installation note 2), which will be on your HARD DRIVE where you manually copied it. Then double clicking on the Run icon.

Descriptions of the Program:

The program consists of three components; **Create Exam, Sit Exam and Edit Exam**.

a. Create Exam Component:

In the Create Exam component, the examiner constructs the exam using any number of boxes up to 12. The exam's name and the file name (where the results will be stored) should be specified in this component. Three icons are found at the bottom of the component. These are **Create Intro** (to write an introductory paragraphs for the exam as a whole or for each individual question), **Save Exam** (to save the exam in the hardware) and **Next Button** (for writing another question).

Grid questions can go beyond selections to ask pupils to put their selections in a logical order. The program allows the assessor to ask these questions and score them very easily. It is required from the assessor to determine in the Create Exam component the type of questions he wants to ask (i.e. selections only or order only or a combination of selections and order).

Sometimes the assessor would like to give more weight to the selection part than to the logical order part or vice versa. In this case, the assessor should determine which part he would like to give more weight, while he is writing the exam, in the space provided in

Create Exam Component. The spaces are called Selective Multiplier (for the selection part) and Logical Multiplier (for logical order part). For example, suppose that the assessor would like to give more weight to the selection part, then he should type the larger multiplying number in the 'Selective Multiplier' and smaller multiplying number in the 'Logical Multiplier'. The same can be applied in case that the assessor would like to give more weight to the logical order part and less weight to the selection part. However, if the assessor would like to give the same weight for both parts, then he should type the same number (1) in both multipliers.

b. Sit Exam Component:

This component is designed where students can practise the exam. Initially, the program asks the student to enter his name so that the program can keep a record and store the results of student's answers. Then, an introductory paragraph(s) will appear instructing the student how to answer the exam.

The following steps show how the pupil answers the exam:

- Student should select the number of question he wants to answer (**Q1, Q2, Q3,Q4…….etc.**) from the top bar.
- Double click on the options he would like to include in his answer will go automatically to the place where the answers should be put.
- If the student changes his mind or makes a mistakes, there is an 'UNDO' facility which enables him to return and make changes.
- If the question asks the student to order his selection logically, then the student should select the 'ORDER' icon after he makes the selections. This icon enables him to shuffle the options to reach a best order.
- If the student wants to finish the exam he should click on 'FINISH' icon. After he does so, the program will ask him a question to make sure that student really wants to end the exam.

c. Edit Exam Component:

This component allows the assessor to edit in the exam he already has in the program. Sometimes, the assessor, for some reasons, would like to add or remove some questions from the existing exam. He can do this very easily by clicking on this component and do the editing. After all the editing is done the assessor should save the exam.

Reading the Results:

To read the results, go to the file where you stored the results on your HARD DRIVE (this file should be specified in the Create Exam component). Then, double clicking on this file will automatically enable you to see the results. The results show the scores that each pupil gets in each question. The assessor can manually calculate the facility value and discrimination index for each question to identify how easy or difficult the question was.

Something to be Aware of:

There are some things that the assessor should be aware of.

1. The assessor can construct many questions based on one grid. However, if he wants to do so he needs to press Next Button in the **Create Exam** component. After pressing this button, the same grid will appear. What you need to do is to:

- Write the new stem in the space provided.
- Put the new correct answers in the space provided.
- Specify the type of the question (selection only, order only or a combination of the two).

2. In terms of question like (***Select boxes a, b, c, d, and e, then order them logically according to ....***) the assessor should remember that in the Create Exam Component he will see numbers instead of letters. However, these numbers were designed to be equal to the letters (i.e. 1 = a, 2 = b, 3 = c, 4 = d, 5 = e, 6 = f ....etc.). Therefore, the assessor still can write questions like this. The only thing he needs to do is to write the numbers (in the Create Exam component) of the correct boxes in the place where the correct answers should be placed so that the computer can identify the correct answers.

3. Unfortunately, the assessor will not be able to write questions based on formulae or equations as in physics and chemistry. Nevertheless, if he wants to ask questions like this then he should converts them into words, if possible.

Appendix (5.2): The Testing Material to Check the Program

Grid Questions in Computer

(Standard Grade Chemistry)

This revision exercise is based on the grid provided for each question. You need to select appropriate letter(s) for each sub - question from the grid provided for each question.

Question One:

A) Butene	B) Propane	C) Pentene
D) Hexane	E) Hexene	F) Octane
G) Pentane	H) Propene	I) Butane

1.a Which box(es) contain(s) hydrocarbons which are Alkanes?

____

1.b Which box contains the hydrocarbon which is used in gas cylinders for homes?

____

Question Two:

A) Iodine	B) Bromine	C) Calcium
D) Mercury	E) Helium	F) Sodium
G) Chlorine	H) Argon	I) Magnesium

2.a Which box(es) contain(s) elements which conduct electricity at room temperature?

____

2.b Which box contains an unreactive gas?

____

2.c.i Which box(es) contain(s) the halogens?

____

c.ii Order your selection in **2.c.i** according to their physical states at room temperature (i.e. gas to solid)?

____

2.d Which box contains the alkaline earth metal?

____

2.e Which box(es) contain(s) the alkali metal(s)?

____

2.f Which box(es) contain(s) the non - metals?

____

Question Three:

A) Calcium Chloride	**B)** Chlorine (Atomic Number = 17)	**C)** Potassium Sulphate
D) Magnesium (Atomic Number = 12)	**E)** Oxygen (Atomic Number = 8)	**F)** Methane
G) Fluorine (Atomic Number = 9)	**H)** Ammonia	**I)** Calcium (Atomic Number = 20)

3.a Which box(es) contain(s) the ionic compounds?

____

3.b Which box(es) contain(s) the elements which could form an ionic compounds with the element in box **D**?

____

3.c Which box contains the element which has an electron arrangement of 2, 8, 2?

____

3.d Which box(es) contain(s) the elements which could form a covalent compounds with carbon?

____

3.e Which box(es) contain(s) the elements which can gain an electron(s) to became negatively charged?

____

Question Four:

A) Sodium Hydroxide	B) Sulphuric Acid	C) Potassium Chloride
D) Sodium Chloride	E) Nitric Acid	F) Potassium Hydroxide
G) Carbonic Acid	H) Water	I) Hydrochloric Acid

4.a Which box(es) contain(s) the compounds that produce hydrogen ions when they dissolve in water?

____

4.b Which box(es) contain(s) the substance(s) which are alkalis?

____

4.c Which box(es) contain(s) the compound(s) that can produce a solution with a pH value of 11 when dissolved in water?

____

4.d Which box(es) contain(s) the compound(s) of the pH value of 7?

____

4.e Which box(es) contain(s) the substance(s) which are salts?

____

4.f Which box(es) contain(s) the pair of substances, when mixed in water, would produce sodium sulphate?

____

Question Five:

A) Methane	**B)** Sulphuric Acid	**C)** Nitrogen
D) Oxygen	**E)** Calcium Chloride	**F)** Ammonia
G) Water	**H)** Carbon Dioxide	**I)** Hydrogen

5.a Which box(es) contain(s) the elements of valency number of 5?

____ ____ ____ ____ ____ ____ ____ ____ ____

5.b Which box(es) contain(s) the molecules with double bonds ?

____ ____ ____ ____ ____ ____ ____ ____ ____

Question Six:

A) increase in particle size	B) dissolving sugar in a cup of tea	C) decrease in concentration
D) decrease in particle size	E) gas burning in a cooker ring	F) decrease in temperature
G) bubbles forming in a cake as it rises	H) an epoxy resin glue setting	I) boiling water in an electric kettle

6.a Which box(es) contain(s) an example of a chemical change?

____

6.b Which box contains the factor that responsible for the following observation: Milk goes sour more slowly in a refrigerator than an identical carton left on a work surface in a kitchen?

____

6.c Which box contains the factor that responsible for the fact that 1 kg of wood shavings burns faster than a block of wood of equal mass?

____

6.d Which box contains the factor that responsible for the following observation: Battery acid and acid rain both contain sulphuric acid but iron nails corrode more slowly in acid rain water than in battery acid"?

____

6.e Which box(es) contain(s) an example of a physical change?

____

Question Seven:

A) Oxygen	**B)** Petroleum gas (less than 40 degree centigrade)	**C)** Heavy gas oil (300 - 350 degree centigrade)
D) Carbon Dioxide	**E)** Residue (more than 350 degree centigrade)	**F)** Naphtha (40 - 180 degree centigrade)
G) Kerosene (180 - 250 degree centigrade)	**H)** Nitrogen	**I)** Carbon Monoxide

7.a.i Which box(es) contain(s) the gases which make up the air?

____

a.ii Order your selections in **7.a.i** according to their percentage in air (i.e. large percentage to low percentage)?

____

7.b Which box(es) contain(s) the hydrocarbons which catch fire more easily than those in kerosene?

____

7.c Which box(es) contain(s) larger hydrocarbon molecules than those in light gas oil?

____

Grid Questions in Computer

(Standard Grade Biology)

The revision exercise is based on the grid provided for each question. You need to select appropriate letter(s) for each sub - question from the grid provided for each question.

Question One:

A) Nucleus	B) Osmosis	C) Cytoplasm
D) Diffusion	E) Vacuole	F) Cell Membrane
G) Paramecium	H) Butterfly	I) Amoeba

1.a Which box contains the structure which is found only in a plant cell?

___ ___ ___ ___ ___ ___ ___ ___ ___

1.b Which box(es) contain(s) the structures which are found in both animal and plant cells?

___ ___ ___ ___ ___ ___ ___ ___ ___

1.c Which box contains the word which has the following meaning. "The movement of the molecules of a substance from a region of high concentration of that substance to a region of low concentration until the concentration becomes equal throughout"?

___ ___ ___ ___ ___ ___ ___ ___ ___

1.d Which box(es) contain(s) an example of a unicellular animal?

___ ___ ___ ___ ___ ___ ___ ___ ___

1.e Which box contains the structure which contains the cell sap?

____

1.f Which box contains the structure that controls the passage of a substance into and out of a cell?

____

1.g Which box contains the structure that is the site of the chemical reaction essential for life?

____

Question Two:

A) New cell wall forming	**B)** Reproduction	**C)** Cytoplasm dividing
D) Two daughter cells formed	**E)** Nucleus dividing into two	**F)** Growth
G) The cell is about to divide	**H)** Chromosome	**I)** Centromere

2.a.i Select from the grid pieces that show events that occur during the cell division in an animal cell??

___ ___ ___ ___ ___ ___ ___ ___ ___

a.ii Now place the pieces you have chosen in the correct order, from start to finish of cell division?

___ ___ ___ ___ ___ ___ ___ ___ ___

2.b Which box contains the structure which joins two chromatids together?

___ ___ ___ ___ ___ ___ ___ ___ ___

2.c Which box contains the function of mitosis in a unicellular organism?

___ ___ ___ ___ ___ ___ ___ ___ ___

Question Three:

A) Dogs, cats, tigers and lions	B) The sum total of all the populations of plants, animals and micro-organisms living together in an ecosystem	C) A group of organisms of the same species.
D) Daffodil, daisy and tulip	E) The place where an organism live	F) A relationship where one organism feeds on the previous one and in turn provides food for the next one in the series
G) Human beings	H) The balanced interaction between the members of a community and their physical habitats	I) A rectangular - shaped sampling unit of known area

3.a Which box(es) contain(s) the description of the word **'population'**?

____

3.b Which box contains examples of the 'carnivores'?

____

3.c Which box(es) contain(s) the description of the word 'ecosystem'?

____

3.d Which box contains an example of 'omnivore'?

____

3.e Which box(es) contain(s) the description of the 'food chain'?

____

Question Four:

A) Owl	**B)** Oak Leaf	**C)** Shrew
D) Greenfly	**E)** Weasel	**F)** Fox
G) Caterpillar	**H)** Bank Vole	**I)** Acorn

4.a Which box(es) contain(s) examples of the primary consumers?

____

4.b Which box(es) contain(s) examples of the producers?

____

4.c Order the organisms in boxes **A, B, C and D** in such a way that they make a food chain?

____

Question Five:

A) Nutrients in living organisms	B) Nitrogen in plants	C) Nitrogen in nitrites in soil
D) Nutrients in environment available for use	E) Nitrogen in dead bodies and wastes	F) Nutrients in dead bodies and wastes
G) Nitrogen in animals	H) Nitrogen in ammonium compounds in soil	I) The nutrients are absorbed by living things

5.a.i Select from the grid the pieces relevant to the nitrogen cycle?

____

a.ii Order your selections in **5.a.i** logically starting from box **G** and give the sequence in the nitrogen cycle?

____

5.b Which box contains the step in which the roots absorb the nitrogen?

____

5.c Which box contains the step in which the nutrients are decomposed by bacteria and fungi?

____

5.d Which box(es) contain(s) the steps in which the nitrification by bacteria is taking place?

____

Question Six:

A) Fresh Water	**B)** Sea	**C)** Oil
D) Land	**E)** Decomposition by micro-organisms	**F)** Air
G) Sulphuer Dioxide	**H)** Recycling, burying and burning	**I)** Radioactive Waste

6.a Which box contains the part of environment which has homes and motor vehicles as the main source of pollution?

6.b Which box contains the substance which causes the acid rain?

6.c Which box contains the part of environment that box C is an example of pollutant?

6.d Which box contains the part of environment in which recycling, burying and burning are the possible control measure?

6.e Which box contains the example of pollutant which the nuclear power stations are the main source?

Question Seven:

A) Sepal	**B)** Stamen	**C)** Style
D) Ovary	**E)** Nectary	**F)** Petal
G) Anther	**H)** Embryo Sac	**I)** Stigma

7.a Which box contains the part of a flower which has bright colour to attract insects?

____

7.b Which box contains the female part containing ovules of the flower and working to the centre?

____

7.c Order boxes **A, B, D, E, and F** logically starting from the outer part of the flower?

____

7.d Which box contains the part which is found in the stamen?

____

7.e Which box contains the region where the pollen grains become attached?

____

APPENDIX (6): THE RELATIONSHIP BETWEEN PUPILS' KNOWLEDGE STRUCTURES AND MULTIPLE CHOICE TEST STRUCTURE

Appendix (6.1): Chemistry Word Association Test

Note:

The Word Association Test was presented to pupils as an A5 booklet.

INSTRUCTIONS

"When you hear or see a word, it often makes you think of other words. In this study we would like to find out what other words are brought to your mind by some words in chemistry.

On each page you will find a key word written many times. Say the word to yourself, and then as quickly as possible write the first word comes to your mind in space Number 1. And then continue to write in other spaces other associated words come to your mind.

Continue in this way until your told to turn the next page. There are no right answers. Write as quickly as possible since you are allowed only **one minute (1 MINUTE)** for each page.

Thank you very much.

An Example:

CHEMICAL REACTION

CHEMICAL REACTION	1. ELEMENT
CHEMICAL REACTION	2. COMPOUND
CHEMICAL REACTION	3. PRODUCTS
CHEMICAL REACTION	4. REACTANTS
CHEMICAL REACTION	5. EQUATION
CHEMICAL REACTION	6. TEMPERATURE
CHEMICAL REACTION	7. BALANCE
CHEMICAL REACTION	8. MOLECULES
CHEMICAL REACTION	9. ELECTRON
CHEMICAL REACTION	10. ATOM

Write as many words as come to your mind in the time available.

PERIODIC TABLE

PERIODIC TABLE	1........................
PERIODIC TABLE	2........................
PERIODIC TABLE	3........................
PERIODIC TABLE	4........................
PERIODIC TABLE	5........................
PERIODIC TABLE	6........................
PERIODIC TABLE	7........................
PERIODIC TABLE	8........................
PERIODIC TABLE	9........................
PERIODIC TABLE	10........................

ELEMENT

ELEMENT	1......................
ELEMENT	2......................
ELEMENT	3......................
ELEMENT	4......................
ELEMENT	5......................
ELEMENT	6......................
ELEMENT	7......................
ELEMENT	8......................
ELEMENT	9......................
ELEMENT	10......................

Appendix (6.2): The Frequency Table and Numerical Codes for some Responses in the Chemistry Word Association Test

Stimulus Words: 1' = Periodic Table 2' = Elements 3' = Compounds 4' = Atoms 5' = Metals 6' = Non Metals 7' = Molecules 8' = Matter

	Stimulus Words							
Responses	1'	2'	3'	4'	5'	6'	7'	8'
1' Periodic Table		10	5	2	3	1		1
2' Elements	36		22	15	2	6	7	4
3' Compounds	6	21		15		1	11	2
4' Atoms	10	15	24		5	2	43	5
5' Metals	13	6	3	3		3	1	3
6' Non Metals								
7' Molecules	3	11	16	37	1	1		4
8' Matter								
9 Aluminum	9	5	1		31	5		
10 Air	2	6	3	8		8	10	3
11 Bend					12	6		1
12 Bomb			1	25				
13 Bricks		1	1	1		11		
14 Balance	1	2	3	2				
15 Bronze	6	5	1		11			
16 Chemical	16	16	16	9	1	1	4	
17 Chemical Reaction		5	7	8			2	
18 Cars					17			
19 Copper	17	9			32	9		1
20 Cloth				1		6		
21 Circles		1	1	8			5	
22 Cup					2	1		
23 Chair				2	7	2	1	
24 Chemistry	4	3	4	2			2	1
25 Dogs			12				1	
26 Electron		1	3	3	1	1	2	2

	Stimulus Words							
Responses	**1'**	**2'**	**3'**	**4'**	**5'**	**6'**	**7'**	**8'**
27 Explosion		2	1	6				
28 Everything				4			4	2
29 Fire		16	1	3	1	2	1	
30 Gas	29	23	13	16		24	13	19
31 Glass		1				25	1	
32 Gold	23	10	2	1	28	8		
33 Hard	1	1		1	37	3	1	2
34 Hydrogen	17	11	3	2		3		
35 Heat			1		2		2	2
36 Helium	8	4	1	2		2		1
37 Iron	7	7	3		34	5		
38 Letters	26	4	1	1				
39 Light	1	2		2	3	3		
40 Lead	6	2	1		7	1		1
41 Liquid	20	10	12	12	6	19	11	16
42 Mixture		2	7	1	1			
43 Mercury	6	1			8	2		
44 Numbers	17							
45 Nickel	2	1	1		3			
46 Nitrogen	7	1	2	1		3		1
47 Oxygen	20	12	4	6		4	1	2
48 Objects		2	2	1	2		1	
49 Paper	1			1	1	33	1	1
50 Plastic			1	1	3	73		
51 Products	3	1	3	1				2
52 Particles	2	5	4	10	1		16	
53 Rubber					1	21		
54 Shine		1			19	1		
55 Science	19	7	6	5	1		9	5

	Stimulus Words							
Responses	**1'**	**2'**	**3'**	**4'**	**5'**	**6'**	**7'**	**8'**
56 Solid	13	10	17	15	22	2	10	16
57 Small		2	2	17			23	5
58 Steel	2	1	2	1	28	3		
59 Silver	17	7	3	1	47	6		1
60 Soft					5	18	1	1
61 States								14
62 Tiny		1		14			21	2
63 Table	8	1	1	4	2	1		
64 Tin	4	1	1		20			1
65 Thing		5	5	3	2	1	2	2
66 Wind		20				1		
67 Water	1	25	7	7	1	21	12	8
68 Wood	1	2			2	63	4	
69 Words	6	2	4				1	3
70 Zinc	8	3	1		14	2		

Appendix (6.3): Chemistry Multiple Choice Test

Appendix 6.3

REVISION EXERCISES

S1 STARTING SCIENCE (CHEMISTRY SECTION)

Name:---

INSTRUCTION

1. This test is for practice only. The result will not affect your academic work or exam in any way.
2. All questions can be answered by drawing a circle around the appropriate letter in each question.
3. All questions are printed on both sides of the paper.
4. Please attempt as many questions as you can.
5. Do not open the test paper until your teacher tells you.

1. Which one of the following elements is invisible at room temperature?

a. zinc.

b. gold.

c. hydrogen.

d. tin.

2. All matter is made up of tiny particles called..................

a. atoms and elements.

b. atoms and molecules.

c. molecules and compounds.

d. molecules and elements.

3. Which one of the following elements has the chemical symbol of 'Ag'?

a. tin.

b. copper.

c. silver.

d. oxygen.

4. Which one of the following elements helps things to burn?

a. oxygen.

b. hydrogen.

c. copper.

d. aluminum.

5. Which one of the following is an element which is not a metal?

a. tin.

b. copper.

c. oxygen.

d. paper.

6. Which one of the following is a chemical element?

a. fire.

b. water.

c. wind.

d. iron.

7. In which of the following substances the molecules are arranged in rows?

a. air.

b. water.

c. steel.

d. oxygen.

8. "It is an element which burns well. It can explode when it is lit". Which one of the following has the above description?

a. hydrogen.

b. oxygen.

c. gold.

d. copper.

9. Which of the following pairs of substance are elements which are in a solid state at room temperature?

a. oxygen and hydrogen.

b. copper and silver.

c. water and wind.

d. glass and plastic.

10. Which one of the following is a compound?

a. air.

b. tin.

c. oxygen.

d. water.

11. Which one of the following is the smallest among the others?

a. molecules.

b. compounds.

c. atoms.

d. elements.

12. In which of the following are the molecules far apart from each other?

a. wood.

b. air.

c. water.

d. glass.

APPENDIX (7): PUPILS' PERFORMANCES IN WORD ASSOCIATION TEST AND THEIR ACHIEVEMENTS IN MULTPLE CHOICE TEST

Appendix (7.1): Biology Word Association Test

Note:

The Word Association Test was presented to pupils as an A5 booklet.

INSTRUCTIONS

"When you hear or see a word, it often makes you think of other words. In this study we would like to find out what other words are brought to your mind by some words in biology.

On each page you will find a key word written many times. Say the word to yourself, and then as quickly as possible write the first word comes to your mind in space Number 1. And then continue to write in other spaces other associated words come to your mind.

Continue in this way until your told to turn the next page. There are no right answers. Write as quickly as possible since you are allowed only **one minute (1 MINUTE)** for each page.

Thank you very much.

An Example:

MEAT

MEAT	1. BEEF
MEAT	2. FAT
MEAT	3. COOK
MEAT	4. ENERGY
MEAT	5. PORK
MEAT	6. BACON
MEAT	7. PROTEIN
MEAT	8. LAMB
MEAT	9. MINCE
MEAT	10. FOOD

Write as many words as come to your mind in the time available.

LIVING THINGS

LIVING THINGS	1........................
LIVING THINGS	2........................
LIVING THINGS	3........................
LIVING THINGS	4........................
LIVING THINGS	5........................
LIVING THINGS	6........................
LIVING THINGS	7........................
LIVING THINGS	8........................
LIVING THINGS	9........................
LIVING THINGS	10........................

ANIMAL

ANIMAL	1........................
ANIMAL	2........................
ANIMAL	3........................
ANIMAL	4........................
ANIMAL	5........................
ANIMAL	6........................
ANIMAL	7........................
ANIMAL	8........................
ANIMAL	9........................
ANIMAL	10........................

Appendix (7.2): The Frequency Table and Numerical Codes for some Responses in the Biology Word Association Test

Stimulus Words: 1' = Living Things 2' = Animals 3' = Plants

4' = Photosynthesis 5' = Energy 6' = Carbon Dioxide 7' = Water 8' = Oxygen

	Stimulus Words							
Responses	**1'**	**2'**	**3'**	**4'**	**5'**	**6'**	**7'**	**8'**
1' Living Things			3					2
2' Animals	58		2	5	9	3	6	35
3' Plants	64			62	11	26	10	38
4' Photosynthesis			6			9		8
5' Energy		2	2	11		3	5	5
6' Carbon Dioxide	1	1	4	14			1	6
7' Water	1	2	23	36	13	6		8
8' Oxygen	5	1	9	32	5	22	4	
9 Air	3			8	1	34		59
10 Alive	4	1	2		2		1	3
11 Birds	20	21		2				
12 Breathing	9	3		1	4	30		67
13 Biology	1	1	2	12	1	5		2
14 Chlorophyll			11	6				
15 Cows	14	30						1
16 Cats	23	57			1			1
17 Bear		11						
18 CO_2	1		8	16	1	20	1	1
19 Clean							26	3
20 Chemistry						21	1	3
21 Clear							19	1
22 Cold					1	1	24	1
23 Dogs	31	51						2
24 Drinking					14	1	50	
25 Daisy			21					
26 Daffodil			23					

	Stimulus Words							
Responses	**1'**	**2'**	**3'**	**4'**	**5'**	**6'**	**7'**	**8'**
27 Exercises					10			
28 Elephant	3	16						
29 Food	5	8	8	16	47	1	3	2
30 Fish	24	20		1	1		15	4
31 Flower	3		26	4		1	1	2
32 Fitness			1		19			
33 Fox		15		2				
34 Gas				3	1	38	1	20
35 Green			34	13	1	1		1
36 Grass	3	2	8	1	2			3
37 Gerbil	1	14						
38 Humans	65	10		2	15	6	4	25
39 Horses	10	28						
40 Heat			1	5	8	1	3	1
41 Health		2			15	1	5	5
42 Hamsters	4	30						
43 Hot			1		1		16	1
44 Insects	30	1	8		2		1	2
45 Leaves	2		22	10		3	1	1
46 Living	1	5	9	4	6	2	4	15
47 Lungs						3		13
48 Life	2	1	2	5	3	1	5	16
49 Leaf		1	9	3				
50 Light			7	34	17	2	1	1
51 Meat	1	9			7			
52 Mammals	15	5				1		2
53 Mouse	3	16					1	1
54 Movement	5	1			10			

	Stimulus Words							
Responses	1'	2'	3'	4'	5'	6'	7'	8'
55 People	19	1	1		5	5	2	9
56 Pansy			18					
57 Pigs	7	31						
58 Protein				1	12		1	
59 Rabbits	8	20			1			
60 Rats	3	12						
61 Roses			36					
62 Roots			12	2				1
63 Rain	1					1	16	
64 Running					29	1	3	
65 River							17	
66 Sheep	8	19		1				
67 Sun			5	22	18	1		1
68 Sea	2						20	
69 Soil	1		15	8				1
70 Swimming					4		15	
71 Small	2	3	4					
72 Starch			4	17				
73 Spiders	8	5	1	1				
74 Snake	5	9		2			1	
75 Tapes							22	
76 Trees	20		21	1		11	1	19
77 Tulips			20					
78 Walking	1				10			

Appendix (7.3): Biology Multiple Choice Test

REVISION EXERCISES

S2 STARTING SCIENCE (BIOLOGY SECTION)

Name:---

INSTRUCTION

1. This test is for practice only. The result will not affect your academic work or exam in any way.

2. All questions can be answered by drawing a circle around the appropriate letter in each question.

3. All questions are printed on both sides of the paper.

4. Please attempt as many questions as you can.

5. Do not open the test paper until your teacher tells you.

1. When carbohydrate burns, it releases the gases?

a. oxygen and carbon dioxide.

b. carbon dioxide and water vapour.

c. nitrogen and water vapour.

d. oxygen and water vapour.

2. Which one of the following is produced in the process of photosynthesis?

a. proteins and oxygen.

b. fats and carbon dioxide.

c. starch and oxygen.

d. vitamins and carbon dioxide.

3. Which one of the following correctly lists what plants need to make food?

a. carbon dioxide, starch, chlorophyll.

b. carbon dioxide, chlorophyll, sunlight.

c. oxygen, starch, sunlight.

d. oxygen, sunlight, water.

4. Iodine is used to test for...................

a. starch.

b. protein.

c. minerals.

d. vitamins.

5. The main source of energy for the Earth is............

a. the sun.

b. the moon.

c. electricity.

d. plants.

6. What colour is observed in a positive starch test?

a. red/brown.

b. blue/red.

c. blue/black.

d. red/black.

7. Which one of the following plants can live for long time without rainfall?

a. buttercup.

b. cacti.

c. daisy.

d. daffodil.

8. Which gas is released from leaves during the process of photosynthesis?

a. carbon dioxide.

b. hydrogen.

c. nitrogen.

d. oxygen.

9. Which one of the following gases makes up most of Earth's Atmosphere?

a. oxygen.

b. nitrogen.

c. carbon dioxide.

d. hydrogen.

10. A 'primary consumer' is ....

a. an animal which eats another animal.

b. an animal which eats the green plants.

c. a plant which consumes bacteria.

d. a plant which builds its own food.

11. Burning fuel uses up.....................

a. oxygen and produces nitrogen.

b. oxygen and produces carbon dioxide.

c. nitrogen and produces carbon dioxide.

d. nitrogen and produces hydrogen.

12. Which one of the following gases changes the colour of lime water?

a. oxygen.

b. nitrogen.

c. carbon dioxide.

d. iodine.